Vol. VI
Part II

The British Journal of Statistical Psychology

November,
1953

THE RELATION OF THE TERMAN-MERRILL VOCABULARY TEST TO MENTAL AGE IN A SAMPLE OF ENGLISH CHILDREN

By M. I. DUNSDON and J. A. FRASER ROBERTS

Burden Mental Research Department, Stoke Park Colony, Bristol,
and Research Department, Royal Eastern Counties Institution, Colchester

I. *Introduction.* II. *Material.* III. *The Correlation between Number of Words and Mental Age at Fixed Chronological Age.* IV. *The Association between Number of Words and Mental Age.* V. *The Association between Number of Words and Chronological Age.* VI. *Vocabulary Norms for the Mental Age-Levels of the Terman-Merril Scale.* VII. *A Sex Difference.* VIII. *Summary.*

I. INTRODUCTION

The surveys in which we have been interested, including those now in progress, necessitate the testing of large numbers of subjects. For example, we are engaged in a study of low-grade mental deficiency, in which part of the plan is to test the brothers and sisters of the subjects. The most important initial question is whether or not the sibs, other than the few who are also low-grade defectives, are on the average of lower intelligence than the general population : in such work tests are used for comparing relatively large groups, and the interpretation of individual results is not required. The limiting factor is the amount of testing time needed ; and the use of individual tests taking an hour or more to administer imposes a severe restriction on what can be done. As an alternative a vocabulary test is an obvious choice. Since Kirkpatrick ((3), 1891 ; (4), 1907) first suggested a vocabulary test for children, much work has been done on vocabularies, and their merits have been widely recognized. Perhaps the best-known version is that introduced by Terman and Childs ((10), 1912) into their revision of the Binet Scale. Hence it seems unnecessary to do more than summarize their suitability for purposes such as ours, and to indicate briefly the work in progress to which this paper is an introduction.

1. In surveys of this kind, completeness of ascertainment, or something very near it, is essential. There is probably no other type of measurement with which it is easier to get a biased sample than those furnished by intelligence-test scores. Outside the limits of school age it is almost impossible to secure a complete sample, so that the age of the subjects must be limited to 5–15 years. But in a survey such as that mentioned all subjects within those limits of age are needed, for the amount of information increases rapidly with the size of the sibship. Hence group-tests are of no use. Vocabulary provides an individual test which is essentially the same test over the whole of the required span of age.

2. It is known that the reliability of vocabulary tests is in general quite high. As a rule it appears to average about ·90.

3. Validity too is very high. In the sample described in this paper the partial correlation between number of words and mental age at fixed chronological age is ·84, a figure very similar to those found by previous investigators.

4. It seems likely that practice effect is minimal. The effect of mere repetition of the same vocabulary is apparently nil. McMeeken ((5), 1939) describes a retest with the Stanford (1916) Scale

of 140 children after the lapse of one day. The mean increase of I.Q. was 2·46 points ; but she states that in vocabulary score no change was observed. We are not aware of any work on the effect of attempts to coach children in the general task of word definition ; but it seems unlikely that this could have any appreciable success. The only factor likely to be serious would be practice in defining the actual words to be used. This can be obviated by providing alternative vocabularies, though of course much heavy work is needed to standardize a new one.

5. Above all, there is the great saving in testing time. A partial correlation of ·84 between M.A. and words defined at fixed C.A. means that 70 per cent. of the information is being obtained in 10 per cent. of the testing time. With a test as applied to an individual the nature of the 70 per cent. and of the residual 30 per cent. raises important problems ; but for the comparison of large groups this need not concern us greatly, though naturally one would be careful to consider the possible associations of so verbal a test with such factors as home background and social class.

With these considerations in mind, one of us (M. I. D.) is administering four vocabularies to a relatively large number of children in the City and County of Bristol, the children being selected by a method which, it is hoped, will yield a good random sample. The four vocabularies are the Terman-Merrill, the Mill Hill Vocabularies A and B, and the vocabulary from the Wechsler Intelligence Scale for Children. We are grateful to Professor L. M. Terman, Mr. J. C. Raven, and Dr. David Wechsler for giving permission for their use.

We have, of course, no intention of standardizing the vocabularies in terms of anything other than themselves, and entirely agree with McNemar ((6), 1942) about the impropriety of attempting to convert vocabulary score into some kind of equivalent in mental age, and thereafter working out an I.Q. Nevertheless, it is valuable to be able to relate one of the vocabularies to a complete scale. We were also anxious to know how far the high average correlation taken over a wide span of age varied at different age-levels and whether it is justifiable to use a vocabulary as a single test down to the age of 5 years. Should there be a sex difference, as indeed we have found in our sample, it is necessary to know whether and to what extent mental age is also involved. Vernon ((12), 1949) has published results based on a sample of English children, showing that their vocabulary scores at most of the relevant mental ages considerably exceeded those required for a pass. For these reasons, and because we had a sample already available, we thought that it might be useful to put on record data for yet another sample of English children.

II. MATERIAL

The main part of the data consists of the results obtained by applying the Terman-Merrill test (Form L) to 450 children ranging in age from 7.0 to 14.11 years. The material was assembled for another purpose in connexion with a recent paper (Dunsdon (2), 1953). The tests were given by Mrs. M. D. Jenner and Mrs. A. E. Sedgley, working at the Burden Mental Research Department of Stoke Park Colony, and by Dr. Margaret A. Mellone and Mrs. E. H. Scarr, working at the Research Department of the Royal Eastern Counties Institution, Colchester. They were applied for a variety of purposes. Many of the children were sibs of mental defectives tested in a survey of which some preliminary account has already been published (Roberts (8), 1952) ; but there were a number of other smaller groups. It was decided to make up a sample with a mean I.Q. of 103 (probably closer to the population mean than 100) and a standard deviation of 17. The limiting factor was the number of persons in the higher (though not the highest) ranges for whom test-results were available. In that range of I.Q. in which there were proportionately fewest results all were used. At other parts of the range the appropriate number was selected by a random process from those available. For the present paper 52 records were added for children of C.A. 5.0–6.11, the mean I.Q. and standard deviation being respectively 96·3 and 15·3. Although no attempt was made to select in relation to other factors, such as sex or chronological age, the sample is satisfactory as far as has been discovered except in one respect. A number of the records showing the highest I.Q.s were drawn from those of children (mostly in their twelfth year) who had secured free places at a grammar school ; hence the children in the age-group 11.0–11.11 are of relatively higher intelligence than those in other parts of the range. The effect of this is mentioned later.

III. THE CORRELATION BETWEEN NUMBER OF WORDS AND MENTAL AGE AT FIXED CHRONOLOGICAL AGE

Using the smallest possible units of grouping, months and single words, the over-all correlation coefficients are as follows : C.A. and M.A., ·8000 ; C.A. and words, ·7349 ; M.A. and words, ·9307. The last figure is very close to Raven's ((7), 1948) correlation of ·926 between Mill Hill Vocabulary score and Terman-Merrill M.A. The relevant figure is the partial correlation of M.A. and number of words at fixed C.A. The influence of chronological age can be appropriately removed by the use of partial correlation as it is a measurement not subject to error. It is true that the regressions are not linear, but as the linear functions account for the great bulk of the associations it is not misleading to use them. Furthermore, the non-linear components in the regressions involving C.A. are almost entirely due to the 11-year age-group containing too many children of high intelligence, as mentioned in the previous section. As regards the association of M.A. with C.A., the expected falling off in the rate of increase of M.A. after the age of 12 is shown by our sample, but is not significant with these numbers. The partial correlation is ·842. A part of the association is spurious because vocabulary is used in scoring M.A. ; but this part cannot be large, as the contribution of vocabulary over the relevant mental ages does not average more than 10 per cent.

The correlations between M.A. and number of words, calculated separately for each year group of C.A., are shown in Table I.

TABLE I. CORRELATION BETWEEN MENTAL AGE AND VOCABULARY AT SUCCESSIVE CHRONOLOGICAL AGES

Chronological Age	Number	Correlation
5.0– 5.11	18	·89
6.0– 6.11	34	·71
7.0– 7.11	40	·60
8.0– 8.11	46	·72
9.0– 9.11	54	·82
10.0–10.11	67	·86
11.0–11.11	67	·83
12.0–12.11	54	·90
13.0–13.11	65	·85
14.0–14.11	57	·89

The figures are closely similar to those of other workers. Terman and Merrill ((11), 1937) state that the correlations between vocabulary and the scale as a whole for single age-groups range from ·65 to ·91 with an average of ·81. McNemar ((6), 1942) gives correlations with the composite scale of ·71, ·83, ·86, and ·83 at ages 8, 11, 14, and 18 respectively.

IV. THE ASSOCIATION BETWEEN NUMBER OF WORDS AND MENTAL AGE

The association of M.A. and number of words can now be examined in more detail, the basic data being shown in Table II. The relation is clearly non-linear. Perhaps some apology is needed for curve-fitting ; but it seems to us that it is helpful to obtain smoothed curves by efficient methods. When the fit proves to be very good, as in this instance, with

TABLE II. BIVARIATE FREQUENCY TABLE FOR MENTAL AGE AND NUMBER OF WORDS

Mental Age	Vocabulary : Number of Words																													Total
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	
19.0–19.11																													1		1
18.0–18.11																									1			1	1		3
17.0–17.11																							2	4	1	1	1		2	1	12
16.0–16.11																		1		2	1	2		4	3	1			1		15
15.0–15.11																	1	1	3	2	4	4	5	4	1	3	1				29
14.0–14.11																1	3	3	8	4	3	5	4	2							33
13.0–13.11															6	3	8	8	7	2	1	1			2	1					39
12.0–12.11											1	1	3	4	9	11	10	9	3	1	4										56
11.0–11.11								1	1	1	1	12	10	15	9	6	5	1	1			1									64
10.0–10.11							1		3	5	12	13	4	7	3					1											49
9.0– 9.11								3	11	12	8	7	5	2	2	1															51
8.0– 8.11							4	2	15	20	8	2	1																		52
7.0– 7.11						1	7	5	13	10	3																				39
6.0– 6.11				1		4	6	8	10		1																				30
5.0– 5.11			3		1	6	9	1																							20
4.0– 4.11		1	1	2	1	2																									7
3.0– 3.11	2																														2
Total	2	1	4	3	2	13	27	20	53	48	34	35	23	28	29	22	27	23	22	12	13	13	11	14	8	6	2	1	5	1	502

entirely non-significant remainders which are close to expectation, then one can feel that the data have been reduced to their simplest form, a form which permits trends to be perceived, unobscured by chance fluctuations. It turns out that biquadratics are needed for both regressions, the equations being :

$$y = -12{\cdot}64 + 6{\cdot}941x - 0{\cdot}9849x^2 + 0{\cdot}06982x^3 - 0{\cdot}001664x^4,$$

where y = number of words and x = M.A. (3.0 — 3.11 = 3, etc., . . .) :

$$x = 2{\cdot}882 + 0{\cdot}2696y + 0{\cdot}2675y^2 - 0{\cdot}02691y^3 + 0{\cdot}0008121y^4,$$

where x = M.A. as before and y = number of words in units of grouping (0 — 1 = 1, 2 — 3 = 2, etc., . . .).

The fitting is shown in Table III and the curves are shown graphically[1] in Fig. 1. The

TABLE III. RELATION OF MENTAL AGE AND VOCABULARY : ANALYSIS OF VARIANCE

A. Regression of Vocabulary on Mental Age

Variation of Vocabulary	Degrees of Freedom	Sum of Squares (words)²	Mean Square (words)²	Variance Ratio	Signi-ficance
Linear regression..	1	13,771·94	13,771·94		++
Addition of quadratic term	1	151·95	151·95	35·19	++
Addition of cubic term	1	3·79	3·79	0·88	—
Addition of biquadratic term	1	58·32	58·32	13·85	++
Remainder between arrays	12	57·02	4·75	1·13	—
Within arrays	485	2,035·47	4·20		
Total	501	16,078·49	32·09		

B. Regression of Mental Age on Vocabulary

Variation of Mental Age	Degrees of Freedom	Sum of Squares (years)²	Mean Square (years)²	Variance Ratio	Signi-ficance
Linear regression..	1	4,356·17	4,356·17		++
Addition of quadratic term	1	63·64	63·64	44·82	++
Addition of cubic term	1	0·32	0·32	0·22	—
Addition of biquadratic term	1	7·46	7·46	5·29	+
Remainder between arrays	10	12·86	1·29	0·91	—
Within arrays	487	688·09	1·41		
Total	501	5,128·54	10·24		

quadratic terms are relatively large in both instances. The biquadratic terms effect a substantial improvement ; with the regression of words on M.A. the residual variance between arrays is cut down to a half and with the other regression by a third.

The form of the curves throws light on the age-group correlations of Table I. The most important feature is the steep rise at the youngest ages, followed by a relative flattening and then by a steep rise once again. The high correlation at age 5, the fall to a minimum at age 7, and then the rise almost to the maximum by age 9 are just what would be anticipated.

[1] We are indebted to Mrs. M. G. Young for drawing the figure.

It is not a matter apparently of vocabulary being a less meaningful test for the youngest children, either necessarily so, or because the lower ranges of the scale are less verbal. Rather, it would seem, among young children the task is such that vocabulary correlates very highly with other mental measurements ; then there is an interval during which the number of words defined becomes rather less meaningful ; finally, by age 9 or 10 vocabulary has regained its close association with other tasks.

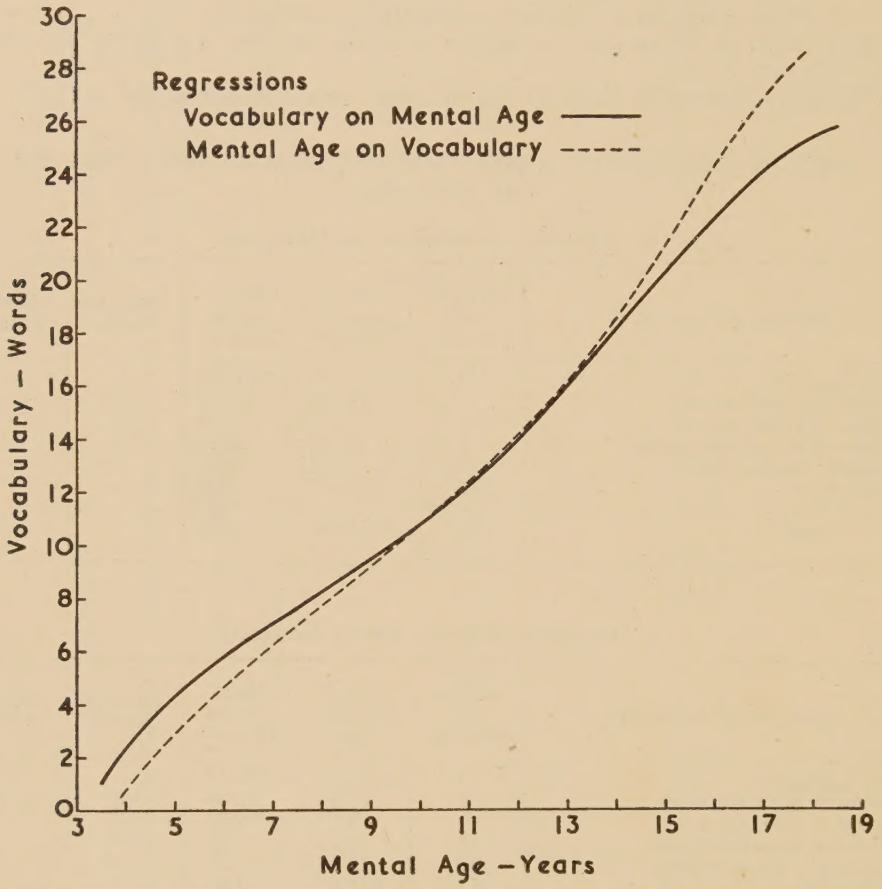

FIG. 1.—Regression of number of words on mental age and regression of mental age on number of words.

It must be recognized, therefore, that from about age 6 to about 8 or 9 validity in terms of the Terman-Merrill Scale is reduced, though not to such an extent that from the practical point of view it is unreasonable to use vocabulary as a single test over the whole range of 5 to 15 years.

V. THE ASSOCIATION BETWEEN NUMBER OF WORDS AND CHRONOLOGICAL AGE

Table IV gives the bivariate frequency table for number of words and chronological age. The association between C.A. and vocabulary is so much looser than that between M.A.

TABLE IV. BIVARIATE FREQUENCY TABLE FOR CHRONOLOGICAL AGE AND NUMBER OF WORDS

Chronological Age	Vocabulary : Number of Words																													Total
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	
14.0–14.11								1	1			3	2		6	4	7	6	5		1	3	3	4	5	1		1	3	1	57
13.0–13.11								1			1	1	1	7	6	5	4	4	10	7	4	1	4	3	1	2	2		1		65
12.0–12.11										5	4	4	1	5	4	3	8	4	1	2	3	2	2	4		1			1		54
11.0–11.11									1	2	6	8	3	4	6	2	5	6	4	2	4	7	2	2	1	2					67
10.0–10.11						1	1	2	7	11	4	8	10	6	4	3	3	2	2	1	1				1						67
9.0– 9.11							4	2	12	9	6	6	4	5	1	3		1						1							54
8.0– 8.11							4	2	13	10	7	3	2	1	2	2															46
7.0– 7.11						4	3	5	12	9	5	2																			40
6.0– 6.11		1	2	2	1	5	11	6	4	1	1																				34
5.0– 5.11	2		2	1	1	3	4	1	3	1																					18
Total	2	1	4	3	2	13	27	20	53	48	34	35	23	28	29	22	27	23	22	12	13	13	11	14	8	6	2	1	5	1	502

and vocabulary that with our numbers there is not much point in examining the form of the regression lines, particularly since in this sample, as already mentioned, the great bulk of the departure from linearity is due to the accidental circumstance that too many clever 11-year-old children were included. More ample data from a better sample will be forthcoming when the main enquiry is completed. The difference in appearance between the two bivariate frequency tables (Tables II and IV) is, however, very striking. The growth in variability of number of words seems largely to keep pace with the growth in the variability of M.A., but in terms of C.A. the variability of the number of words defined increases very markedly with age. It will be necessary to allow for this in the age standardization of the vocabularies.

VI. VOCABULARY NORMS FOR THE MENTAL AGE-LEVELS OF THE TERMAN-MERRILL SCALE

As regards total score there is reasonably good agreement between the norms given by Terman and Merrill and the results obtained with British children. Nevertheless, many individual items do show differences, some being placed too high for British children and some too low (Burt (1), 1948). The standardization of the vocabulary test has been criticized by several writers as being much too easy. Vernon ((12), 1949, p. 184) gives a table of norms, based on a preliminary research, which differ widely from those given in *Measuring Intelligence*. Table V compares the norms derived from the present sample with those of Terman and Merrill and of Vernon.

TABLE V. NUMBER OF WORDS CORRECTLY DEFINED AT MENTAL AGE-LEVELS IN TERMAN-MERRILL SCALE

Mental Age	No. of Words Correctly Defined		
	Terman and Merrill's Sample	Vernon's Sample	Present Sample
6..	5	5	6
8..	8	9	8
10..	11	13	11
12..	14	17	14
14..	16	21	18
15.4 (A.A.)	20	24	21
17.4 (S.A. I)	23	29	25
19.10 (S.A. II)	26	36	—
22.10 (S.A. III)	30	42	—

The expectations based on the present sample are identical with those of Terman and Merrill at M.A. levels 8, 10, and 12. The difference of one word at M.A. 6 hardly means anything as our expectation is only just over 5.5. At M.A. levels 14, A.A., and S.A. I, our figures lie between those of the other series, but are closer to those of Terman and Merrill. It would be unjustifiable with our sample to extrapolate to S.A. II, but our curve suggests a number much nearer 26 than Vernon's 36. It should be mentioned, however, that it is possible that the present sample may be somewhat overweighted with children from lower social groups ; and if this is so it may be that the number of words defined has been slightly reduced in relation to the scale as a whole.

VII. A SEX DIFFERENCE

The present sample shows a small but highly significant sex difference. The means and linear regressions are as follows :

	No.	Mean C.A. (years)	Mean M.A. (years)	Mean No. of words	Increment in words per year of C.A.	Increment in words per year of M.A.
Boys ..	260	10·635	10·881	13·150	1·674	1·681
Girls ..	242	10·706	10·940	12·665	1·547	1·627
Difference ..		−0·071	−0·059	+0·485	+0·127 ± 0·133	+0·054 ± 0·058

Using the regressions to adjust M.A. and C.A. to 10·80 years, the sex differences in favour of boys in number of words defined as at that age are : C.A., 0·616 ± 0·343 ; M.A., 0·575 ± 0·186. Thus, for a given C.A. or M.A., boys define more words than girls, the difference as at 10·80 years being about 0·6 word. The difference is not significant for C.A., but, owing to the close association of number of words and M.A., is highly significant for M.A., being 3·1 times its standard error. It would seem to extend over the whole range of age, and is apparently a matter of the position of the regression lines and not of their slope, though it is possible that the rates of growth might be significantly dissimilar with larger numbers.

It will be noted that it is vocabulary which is slightly sex biased and not mental age as a whole. In another British sample, that of the Scottish Survey ((7), 1949), boys did better on the Terman-Merrill Scale than girls ; but that sample covered only a narrow range of age. In the present sample, with its 10-year span, there is no trace of sex bias with the scale as a whole. Girls are 0·071 year older than boys and their mean M.A. is 0·059 year greater, a difference of one in the first decimal place of I.Q. Agreement between the sexes could hardly be closer.

VIII. SUMMARY

1. Intelligence-tests intended to cover the whole range of school age must be applied individually. For the comparison of large groups vocabulary tests have manifest advantages in respect of reliability, validity, relative absence of practice effect, and, above all, in the amount of information obtained per unit of testing time. Accordingly four vocabularies are being given to school children at Bristol, selected by a method which, it is hoped, will give a good random sample. The present paper presents data from a sample of English children showing the relation of the Terman-Merrill vocabulary test to the scale as a whole.

2. The sample consists of results obtained by applying Form L to 450 children ranging in age from 7.0 to 14.11 years. These were selected from the results available so as to yield a sample with a mean I.Q. of 103 and a standard deviation of 17. To them were added data for 52 children aged 5.0 to 6.11 years, who were slightly lower in I.Q. and slightly less variable.

3. The over-all partial correlation between number of words and M.A. at fixed C.A. is ·842, which is in close agreement with the results of other workers. Thus 70 per cent. of the information is obtained in 10 per cent. of the testing time.

4. The regressions of words on M.A. and of M.A. on words are non-linear. The form of the curves is reflected in lower correlations between number of words and M.A. in the seventh, eighth, and ninth years of life, these being respectively ·71, ·60, and ·72. The other age-group correlations range from ·82 to ·90.

5. The number of words defined at the various M.A. levels of the scale agree almost exactly with those of Terman and Merrill up to 12 years. At older levels the children of the present sample exceed the norms given in *Measuring Intelligence*, but are closer to these norms than to those given by Vernon.

6. In this sample there is no sex difference in performance with the Terman-Merrill Scale as a whole. There is, however, a sex bias in vocabulary score ; for given M.A. boys define more words than girls.

REFERENCES

1. Burt, C. (1948). *Handbook of Tests for Use in Schools* (2nd ed.). London : Staples Press.
2. Dunsdon, M. I. (1953). ' A Comparison of Terman-Merrill Scale Test Responses among Large Samples of Normal, Maladjusted, and Backward Children.' *Jour. Ment. Sci.* (in the press).
3. Kirkpatrick, E. (1891). ' Number of Words in an Ordinary Vocabulary.' *Science*, XVIII, 107–8.
4. Kirkpatrick, E. (1907). ' A Vocabulary Test.' *Pop. Sci. Monthly*, LXX, 157–64.
5. McMeeken, A. M. (1939). *The Intelligence of a Representative Group of Scottish Children.* London : University of London Press.
6. McNemar, Q. (1942). *The Revision of the Stanford-Binet Scale.* Boston : Houghton Mifflin.
7. Raven, J. C. (1948). ' The Comparative Assessment of Intellectual Ability.' *Brit. J. Psych., Gen. Sect.*, XXXIX, 12–19.
8. Roberts, J. A. F. (1952). ' The Genetics of Mental Deficiency.' *Eugen. Rev.*, XLIV, 71–83.
9. Scottish Council for Research in Education (1949). *The Trend of Scottish Intelligence.* London : University of London Press.
10. Terman, L. M., and Childs, H. G. (1912). ' A Tentative Revision and Extension of the Binet-Simon Scale.' *J. Educ. Psychol.*, III, 205–8.
11. Terman, L. M., and Merrill, M. A. (1937). *Measuring Intelligence.* London : Harrap.
12. Vernon, P. E. (1949). *The Measurement of Abilities.* London : University of London Press.

Vol. VI
Part II

The British Journal of Statistical Psychology

November,
1953

CHANGES IN TERMAN-MERRILL I.Q.s WITH DULL CHILDREN

A Test of the Roberts-Mellone Adjustments

By ELIZABETH H. SCARR

Research Department, Royal Eastern Counties Institution, Colchester

I. *Introduction.* II. *The Data.* III. *Comparison of I.Q.s in Test and Retest.* IV. *Comparison between the Observed Mean Changes in I.Q. and those Predicted by the Roberts-Mellone Adjustments.* V. *Summary.*

I. INTRODUCTION

In a recent paper [1] Roberts and Mellone have devised a method of adjusting Terman-Merrill I.Q.s for their lack of independence of chronological age. They describe the fitting of a curve for regression of variance of I.Q. on C.A., obtained by using the test results (Form L) of the 2,000 children in the Terman-Merrill sample whose ages range from 5.0 to 14.11 years. From this curve is derived a table of adjustments, so that the adjusted I.Q.s become comparable at different age levels. In their paper Roberts and Mellone say :

> "An I.Q. of 52 at 5 years 1 month may be expected to rise to a maximum of 54 at 5 years 7 months ; remain constant for about a year ; fall by about 2 points a year to age 8 ; thereafter fall steadily at the rate of 4 points a year to 11 years 1 month ; remain constant to 13 years 1 month ; and finally rise to 49 again. From 14 years 7 months or so the rise becomes suspiciously great. Whether in fact these actual changes will be shown on large samples can only be determined by repeated measurements of the same individuals."

Data for making this practical test have since become available in connexion with retests made on a sample of mentally defective and backward children as part of a study on mental growth at low intelligence levels. As the adjustments are based on variances, their average size in any given group increases with the distance of that group from the mean of the general population. In a group of relatively low I.Q. the adjustments are large ; and consequently the sample thus examined provides suitable material for making the test.

II. THE DATA

The sample consists of a group of backward and mentally defective children attending Special Schools for the educationally subnormal and Occupation Centres in Chelmsford, Colchester, Dagenham, Leytonstone, and Walthamstow, and the Residential Special Schools of the Royal Eastern Counties Institution in Colchester and Cambridge. All the available children in the selected schools, between the ages of 6.0 and 14.11 years, were tested by means of Form L. (Children under the age of 6 years were not included, partly because relatively few find their way into Special Schools and Occupation Centres, and partly because very retarded children are untestable when so young.) So far as possible, the same children were retested, after the lapse of one year. Some difficulty was encountered in obtaining a retest after this fixed

[1] This *Journal*, V (1952), pp. 65–79.

interval, the main obstacles being absence through illness and the intervention of school holidays. About 10 per cent. of the original group could not be retested on a calendar date sufficiently close to that of the original test ; and these were omitted. The complete sample consists of 402 subjects, the age range at first test being 6.0 to 14.11 years. All tests and retests were made by myself.

For testing the adjustments only those children aged 13.11 years or less at the first test could be used, since the correction table for Terman-Merrill I.Q.s gives adjustments to age 14.11 years only. This smaller group consists of 350 subjects [1] whose ages range from 6.0 to 13.11 years at first test. The I.Q.s range from 20 to 85 at the first test, with a mean of 56·97 and a standard deviation of 14·12.

The intervals between first and second tests varied about 365 days in accordance with the following distribution : ± 0–7 days, 207 children ; ± 8–14 days, 79 children ; ± 15–21 days, 41 children ; ± 22–28 days, 23 children.

III. COMPARISON OF I.Q.s IN TEST AND RETEST

TABLE I. FREQUENCY DISTRIBUTION OF DIFFERENCES IN I.Q. BETWEEN FIRST AND SECOND TESTS

2nd I.Q. *minus* 1st I.Q.	1st I.Q. < 40	1st I.Q. > 39			Total
	All C.A.s	C.A. at first test			
		6.0–10.11	11.0–11.11	12.0–13.11	
+ 10				2	2
+ 9	1			2	3
+ 8				2	2
+ 7		1			1
+ 6		3	1	1	5
+ 5	2	4		2	8
+ 4	3	6	1	6	16
+ 3	3	5		12	20
+ 2	1	8	1	11	21
+ 1	7	9	3	8	27
0	12	9	6	8	35
− 1	8	9	3	15	35
− 2	5	16	5	12	38
− 3	7	18	4	10	39
− 4	2	10	4	9	25
− 5	3	9	4	7	23
− 6	2	10	3	2	17
− 7	1	5	2	2	10
− 8		7	2	2	11
− 9		3	1	1	5
− 10		1	1	1	3
− 11					
− 12		1	1		2
− 13		1			1
− 14		1			1
Number	57	136	42	115	350
Mean	− 0·561	− 2·191	− 2·976	− 0·357	− 1·417
Standard Error of Mean	0·403	0·355	0·575	0·367	0·213

[1] The actual number of children was 339, of whom 11 were tested three times. The 11 are too few for separate treatment and the analysis is made in terms of the 350 one-year differences.

Table I shows the frequency distributions of differences between test and retest. (*a*) The mean difference between the two tests, for the whole group of 350 subjects, is $-1{\cdot}417 \pm 0{\cdot}213$; the standard deviation of differences is 3·977. (*b*) The mean difference for the whole group, ignoring sign, is $3{\cdot}320 \pm 0{\cdot}139$, with a standard deviation of 2·604. (*c*) The coefficient of correlation between first and second I.Q. is $+ 0{\cdot}960$.

The grouping of chronological age and I.Q. used in Table I anticipates the findings of the next section. It will be seen that this grouping shows most clearly the trends of agreement and disagreement between the observed changes in I.Q. and those predicted by the Roberts-Mellone adjustments.

IV. COMPARISON BETWEEN THE OBSERVED MEAN CHANGES IN I.Q. AND THOSE PREDICTED BY THE ROBERTS-MELLONE ADJUSTMENTS

The method of finding the predicted change in I.Q. during the year between the two tests was as follows. The I.Q. obtained in Test 1 was located in the correction table, Table VIII (*loc. cit.*, pp. 72–3), opposite chronological age for Test 1, and the adjustment at the head of the column noted. Then the same obtained I.Q. at Test 1 was again located in the body of the table, this time opposite chronological age for Test 2 (i.e., one year later), and again the appropriate adjustment at the head of the column was noted. The difference between these two adjustments gives the predicted rise or fall in I.Q. over the period of the one year between the two tests. For example, for I.Q. 63 at age 10.0 years, the correction is + 5 ; for the same I.Q. at age 11.0 years it is + 7. Thus the predicted fall in I.Q. over the year is 2 points. We have, then, for each subject the observed change in I.Q. and the predicted change in I.Q. By comparison of the two the applicability of the adjustments can be tested at the various age and I.Q. levels of the 350 subjects.

The observed and predicted mean changes in I.Q. are shown in Table II.

Theoretical considerations as well as practical experience indicate that there must be a point in intelligence-level, though not necessarily a clearly defined one, below which adjustments are inapplicable. The figures of Table II show that in the present sample an I.Q. of about 40 hits off this point very fairly. In comparing the observed and predicted mean changes, therefore, I.Q.s under 40 have been omitted from the means shown in the right-hand totals column of Table II. This also explains why in Table I a separate frequency distribution of differences is shown for all I.Q.s under 40.

From inspection of Table II it can be seen that the agreement between observed and predicted changes in I.Q. is, on the whole, very good up to the age of 10.11 years at first test. As the predicted change in a child of 10.11 years refers to growth from that age up to 11.11 years, this means that the adjustments are applicable up to a chronological age of about 12 years. For this span of chronological age, 6 to 12 years, the Roberts-Mellone fitted curve for variance of I.Q. on C.A. predicts a substantial fall in I.Q. over the lower ranges of I.Q. This is clearly borne out by the data. During the thirteenth year, the curve predicts a period of stability, during which little change in I.Q. takes place ; the actual results, however, show a continued fall in I.Q. During the fourteenth and fifteenth years, the Roberts-Mellone curve predicts a fairly steep rise in low I.Q. (As the variance decreases, low I.Q.s should be rising.) Instead of this, the data show that the levelling off, which was predicted in the thirteenth year, is in fact taking place in the fourteenth and fifteenth years.

TABLE II. MEAN CHANGES IN I.Q. : OBSERVED CHANGES COMPARED WITH THOSE PREDICTED BY ROBERTS AND MELLONE'S ADJUSTMENTS

C.A. at 1st test	I.Q. at 1st test																	
	< 40			40–49			50–59			60–69			> 69*			Total > 39		
	No.	Obs.	Pred.	No.	Obs.	Pred.	No.	Obs.	Pred.	No.	Obs.	Pred.	No.	Obs.	Pred.	No.	Obs.	Pred.
6.0–6.11	4	+ 1·0	− 3·0	4	− 2·0	− 2·5				3	− 2·0	− 2·3	1	0·0	− 1·0	8	− 1·75	− 2·25
7.0–7.11	5	− 1·4	− 5·0	2	+ 2·0	− 4·0	5	+ 3·0	− 3·2	6	− 3·0	− 2·5	3	− 5·3	− 2·0	16	− 0·94	− 2·81
8.0–8.11	7	+ 1·4	− 4·7	3	− 4·0	− 3·7	8	− 0·6	− 3·1	6	− 3·0	− 2·7	12	− 1·8	− 2·1	29	− 1·97	− 2·66
9.0–9.11	10	− 0·4	− 3·7	1	− 2·0	− 2·0	10	− 1·5	− 2·5	16	− 2·6	− 2·1	12	− 3·3	− 1·7	39	− 2·51	− 2·05
10.0–10.11	5	− 2·4	− 2·2	6	− 1·7	− 1·8	12	− 2·3	− 1·8	17	− 2·8	− 1·3	9	− 3·1	− 1·0	44	− 2·59	− 1·43
11.0–11.11	10	− 0·6	− 0·4	3	− 3·7	− 0·3	16	− 1·9	− 0·1	12	− 3·6	− 0·3	11	− 3·6	− 0·3	42	− 2·98	− 0·21
12.0–12.11	7	− 1·6	+ 1·4	8	+ 0·5	+ 1·1	15	+ 1·2	+ 1·5	24	− 1·3	+ 1·0	11	− 2·7	+ 0·6	58	− 0·67	+ 1·09
13.0–13.11	9	− 0·7	+ 5·9	9	− 0·8	+ 5·0	17	+ 0·9	+ 4·4	21	+ 0·1	+ 3·6	10	− 1·1	+ 2·3	57	− 0·04	+ 3·82

* Includes six I.Q.s of 80 to 85.

TABLE III. SUMMARY OF OBSERVED AND PREDICTED MEAN CHANGES IN I.Q.

I.Q. at 1st test	C.A. at 1st test								
	6.0–10.11			11.0–11.11			12.0–13.11		
	No.	Obs.	Pred.	No.	Obs.	Pred.	No.	Obs.	Pred.
40–49	16	− 1·8	− 2·6	3	− 3·7	− 0·3	17	− 0·2	+ 3·2
50–59	35	− 0·9	− 2·5	16	− 1·9	− 0·1	32	+ 1·0	+ 3·0
60–69	48	− 2·7	− 1·9	12	− 3·6	− 0·3	45	− 0·7	+ 2·2
> 69	37	− 2·9	− 1·7	11	− 3·6	− 0·3	21	− 1·8	+ 1·4
Total	136	− 2·19	− 2·08	42	− 2·98	− 0·21	115	− 0·36	+ 2·44
Std. Error of Obsd. Mean		± 0·355			± 0·575			± 0·367	

Table III, which is a summary of the results shown in Table II, clearly shows these trends.

In their paper, Roberts and Mellone also describe corresponding calculations carried out on test results from a sample of 689 children between the ages of 5.0 and 14.11 years, being the sibs and cousins of a group of mentally defective children. This sample was obtained from areas around Bristol and Colchester. The correspondence with the Terman-Merrill sample was close except in the oldest children. An alternative series of multipliers, based on the English sample, was included for comparison in their Table VIII. It seemed worth while calculating the changes of I.Q. predicted on the basis of this latter sample. It might have been thought that the observed changes in I.Q. would correspond more closely with those predicted on the basis of a sample of English children. This was not, however, found to be so, as is clearly revealed in Table IV. With the solitary exception of ages 9.0 to 9.11 years, there is closer agreement with the changes predicted on the basis of the Terman-Merrill sample.

TABLE IV. COMPARISON OF OBSERVED MEAN CHANGES WITH THOSE PREDICTED ON THE BASIS OF TERMAN-MERRILL AND ENGLISH SAMPLES

C.A. at 1st test	Number	Mean change		
		Observed	Predicted	
			Terman-Merrill Sample	English Sample
6.0– 6.11	8	− 1·75	− 2·25	− 3·13
7.0– 7.11	16	− 0·94	− 2·81	− 3·44
8.0– 8.11	29	− 1·97	− 2·66	− 3·00
9.0– 9.11	39	− 2·51	− 2·05	− 2·13
10.0–10.11	44	− 2·59	− 1·47	− 1·07
11.0–11.11	42	− 2·98	− 0·21	+ 0·26
12.0–12.11	58	− 0·67	+ 1·09	+ 2·36
13.0–13.11	57	− 0·04	+ 3·82	+ 8·04

There remains the question of the effect of practice in the test on the second obtained I.Q. As the interval between Test 1 and Test 2 was one year, and the group consisted of children of subnormal intelligence, it has been assumed that the practice effect was negligible. If it had been appreciable, one would have expected it to manifest itself in the higher I.Q. ranges. The results, however, show no evidence for this.

V. SUMMARY

1. In a recent paper Roberts and Mellone (1952) examined the relation of variance of Terman-Merrill I.Q. to C.A. and devised adjustments intended to correct I.Q.s for their lack of independence of chronological age.

2. The retesting of 350 mentally defective and backward children after the lapse of a year enables a test to be made of the validity of the adjustments. The children ranged in age at first test from 6.0 to 13.11 years, with a mean I.Q. of 57.

3. In children of all ages the results show that the adjustments are inapplicable to I.Q.s below a level of about 40.

4. With I.Q.s of 40 and over the changes predicted by the adjustments are closely realized up to the age of about 12 years.

5. During the thirteenth year the adjustments presuppose a levelling off, whereas in the sample the mean I.Q. continues to fall. During the fourteenth and fifteenth years the adjustments presuppose a rise in I.Q., whereas in the sample there is a levelling off. Thus in this sample, apparently, the curve of variance of I.Q. on C.A. is inflected about a year later than was predicted.[1]

[1] *Acknowledgements.* I am greatly indebted to the Governors of the Royal Eastern Counties Special Schools for the facilities provided and to the head teachers and staff of the schools for their help. I am also greatly indebted to the education authorities and to the head teachers and staff of the Special Schools at Chelmsford, Colchester, Dagenham, Leytonstone, and Walthamstow, and to Dr. Kenneth Cowan and the Supervisors of the Junior Occupation Centres of the County of Essex. My special thanks for help and advice are due to Dr. R. M. Bates, Dr. J. A. Fraser Roberts, and Dr. Margaret A. Mellone.

Vol. VI Part II The British Journal of Statistical Psychology November, 1953

THE SUBMISSIVENESS OF NATIONS

By LEWIS F. RICHARDSON
Hillside House, Kilmun, Dunoon, Argyll, Britain

I. *Introduction.* II. *The Linear Theory of Arms-races.* III. *First Hypothesis about Submissiveness.* IV. *Second Hypothesis about Submissiveness.* V. *Third Hypothesis (for Sides Equal in their Constants).* VI. *Conclusion.*

I. INTRODUCTION

The problem to be discussed in this paper can be stated quite simply as follows. It is well known that a thorough military defeat leaves the vanquished nation for some years in a mood of submission to its victors (examples are the submissive behaviour of Vichy-France to the Germans during 1941 to 1943, or of West Germany to the Western Powers and of East Germany to the Soviet Union during 1945 to 1952). As this effect seems universal, it is not necessary to name the nations ; they may be distinguished as A and B. Now suppose that A and B are both rested from war and are roughly equal in size and in industrial power ; and suppose further that A makes a show of force against B. Does such behaviour induce in B a submissive mood or a retaliatory mood ?

Whatever may be thought of the following treatment, there can be no doubt that some of the questions connected with submissiveness are important. For example, in September, 1952, the North Atlantic Treaty Organization held combined naval manœuvres (exercise " Mainbrace ") off Norway and in the Baltic simultaneously with land manœuvres (exercise " Holdfast ") involving nearly 200,000 troops in Western Europe. The declared intention of N.A.T.O. was to deter any tendency in the Soviet Union towards aggression. On the contrary the Russian newspaper *Pravda* said that exercise " Mainbrace " was itself aggressive (B.B.C., September 13th). Personally I accept N.A.T.O.'s description of its intention ; but I appreciate that *Pravda* may find this intention very difficult to believe, because of the universal tendency to suspect foreign forces unless those of allies—a tendency represented in the following formulæ by the defence coefficients k or l. To be deterred from any tendency towards aggression is here regarded as a mild form of submission.

Professor C. A. Mace asked whether my paper contained any suggestion for research on small groups of people, a type of enquiry now being pursued by many psychologists. Probably it does ; for submission has been occasionally observed wherever two or more individuals meet. For examples : even between wolves (Lorenz (14), 1952, ch. 12) ; or the pecking order in the hen yard (references are given by Landau (11), 1951) ; or tests on individual persons in their relation to the community (Allport (1), 1945, pp. 410–14). W. McDougall (1917 (15), pp. 62–6) took self-subjection to be one of the primary human instincts. It might be rash, however, to expect that by research on a group of, say, 50 people one could predict the behaviour of a nation : for, at least in physics, large aggregates of molecules show some conspicuous forms of behaviour that are not noticeable in small aggregates. For example, I doubt whether anybody, by intensive observation of a drop of water, could have discovered the peculiar law of eddy-diffusion which operates in the sea (Richardson and Stommel (27), 1948).

II. THE LINEAR THEORY OF ARMS-RACES

This theory must first be mentioned, because submissiveness was formulated by a modification of it. From a pamphlet by Bertrand Russell (28) in 1914 I took the idea that much of what is nowadays called defence does not in fact defend, because between

equal groups in peacetime it provokes retaliation. Gregory Bateson ((2), 1936, p. 266) observed competitive boasting in New Guinea and generalized the concept as ' symmetrical schismogenesis,' of which an arms-race is another example. Russell's concept of mutual stimulation was formulated by me as a pair of simultaneous differential equations (1919, revised 1935). In 1938 I hit on a way of testing the formulæ quantitatively in terms of expenditure. The testing was continued in a monograph published a few months before the war (*Generalized Foreign Politics*, *Brit. J. Psychol. Mon. Sup.*, 1939 ; a better title would have been *The Theory of Arms-races*).

The financial facts have been re-examined in much greater detail in a microfilm called *Arms and Insecurity* ((23), 1949), where the technical term ' warfinpersal ' is introduced as a contraction for ' war finance per salary.' The ' warfinpersal ' of any nation is its governmental warlike (alias ' defence ') expenditure divided by the annual earnings of an average sort of citizen, usually those of a semi-skilled engineer. The latest summary, including a glance at the present arms-race, is being printed for *Sankya* (26). The sceptical reader is referred to the aforesaid publications for the facts ; here there is room only for the equations which summarize them. As we pass from the study of individual-differences to that of nations, the vagaries of individuals merge into the quasi mechanical regularities of mass-behaviour.

Let x and y denote the warfinpersals of the opposing sides, t the time, and g, h, α, β, k, l be constants of which α, β, k, l are positive. The hypothesis is that

$$dx/dt = g - \alpha x + ky, \qquad (2/1)$$

$$dy/dt = h - \beta y + lx. \qquad (2/2)$$

In particular, if $g = 0$, $h = 0$, $k = l$, and $\alpha = \beta$, then the solution is

$$x + y = Ae^{(k-\alpha)t},\ x - y = Be^{-(k+\alpha)t}, \qquad (2/3,\ (2/4)$$

where A and B are arbitrary constants. Functions like $x + y$ and $x - y$, which are proportional to a single exponential of the time, are called ' normal co-ordinates.' They lead on well to modifications intended to represent submissiveness : that is why, in the subsequent diagrams, the axes are turned at 45° to those of x and y.

When we are trying to interpret historical facts, we are not given that $\alpha = \beta$ and that $k = l$ exactly, although we may have reasons based on population, on industrialization, and on an assumption about common human nature, for believing that α/β and k/l are not far from unity. In these circumstances the normal co-ordinates would not be simply $x + y$ and $x - y$, but would be $l_1x + m_1y$ and $l_2x - m_2y$, where the l and m are positive weights. Now in an arms-race x is roughly proportional to y, and therefore $x + y$ is roughly proportional to $l_1x + m_1y$ for any not too different weights l_1, m_1. So we can safely study the historical statistics by replacing the unknown $l_1x + m_1y$ by simply $x + y$. But there is no such automatic immunity from error in a replacement of $l_2x - m_2y$ by $x - y$; for they may even have different signs. This difficulty crops up in Sect. V below, where $x - y$ suits the First but not the Second World War.

All the varieties of motion that can be described by linear equations like (2/1) and (2/2), but without any restriction on the signs of α, β, k, l, have long been known (Lagrange, Weierstrass (31), 1875). They were considered, in the context of arms-races, by Richardson ((23), 1949, § 3/2/4) ; but none of them was found to resemble the phenomena of submissiveness, which must therefore require either variable coefficients, a higher order differential equation, an integral equation, or a non-linear description. That was my objection to a linear hypothesis with constant coefficients proposed by Mr. M. R. Horne as a representation of submissiveness ((9), 1951).

III. FIRST HYPOTHESIS ABOUT SUBMISSIVENESS

The statesmen of U.S.A., Britain, France, and U.S.S.R. still appear to pay no attention to mutual stimulation, for they take a pride in accumulating the type of weapons, such as bombing aircraft, or rockets, or submarines, which are most likely to provoke countermeasures. Perhaps they are relying on some unexpressed theory about submissiveness, thus : much force, much submission ; therefore a show of force, some submission. That I call " thinking proportionately." The behaviour of living things is not always so simple. I remember an impertinent schoolboy who was excited and made more rebellious by various mild reproofs and punishments from a schoolmaster in the presence of a class of amused boys, but who was finally suppressed by a much more severe punishment. The behaviour of nations seems to involve a similar reversal.

In 1939 Richardson (21) modified (2/1) and (2/2) by non-linear terms intended to represent submissiveness, thus :

$$dx/dt = g - \alpha x + ky\{1 - \sigma(y - x)\}, \qquad (3/1)$$

$$dy/dt = h - \beta y + lx\{1 - \rho(x - y)\}, \qquad (3/2)$$

where σ and ρ were positive constants representing the tendency to submit.

Application of the First Hypothesis (a) *To Very Unequal Nations.*—Eqn. (3/1) and (3/2) were originally intended to represent a large x-nation overawing a much smaller y-nation. If we leave out g and h, as not of the essence of the matter, and if we neglect y in $x - y$, these equations become

$$dx/dt = -\alpha x + ky\{1 + \sigma x\}, \qquad (3/3)$$

$$dy/dt = -\beta y + lx\{1 - \rho x\}. \qquad (3/4)$$

It is dy/dt that will become negative. The smaller nation is suppressed.

For example in 1948 claims by Guatemala to British Honduras were suppressed by a show of armed force by Britain, without any fighting (K9153, K11511). I am not suggesting that the suppression of a smaller nation is necessarily either just or unjust ; but am merely asserting that it is what usually happens in a world of power politics.

(b) *To Nations Equal as to their Constants.*—In 1951 I noticed that the terms representing submissiveness would have remarkable consequences even if the two opponents were equal in their constants. How this might happen was explained in a letter to *Nature* (1951, ix, 29) headed " Could an arms-race end without fighting ? ", in which arguments about increase or decrease were applied only in circumstances not far removed from the actual. Wishing, however, to see a general integral formula over the whole field, and not finding any, I applied to Dr. M. L. Cartwright, F.R.S., who has made a special study of non-linear differential equations ((4), 1949, (6), 1952). As an apology for the lack of general integral formulæ in the first two theories, I may quote her remark ((5), 1952, p. 88) that " what the pure mathematicians have done shows that the non-linear phenomena are genuinely complicated, and no easily applicable general theory can be expected." As the subject matter is economic or political, 3 per cent. accuracy coupled with a plainly intelligible meaning is to be preferred to infinite accuracy coupled with a complicated formula.

A theory of the submissiveness of equal groups must not be too deterministic : it must not say *which* of them will submit, for that is obviously a chancy affair. Such " stochastic processes " have been the subject of a symposium at the Royal Statistical Society (Moyal, J. E., Bartlett, M. S., and Kendall, D. G. (16), 1949).

By the time an arms-race has gone so far as to be alarming, former grievances and ambitions seem relatively unimportant ; so that we may neglect g and h. On setting also $\alpha = \beta$, $k = l$, $\sigma = \rho$ eqn. (3/1) and (3/2) simplify to :

$$dx/dt = -\alpha x + ky + k\sigma(xy - y^2), \qquad (3/5)$$

$$dy/dt = kx - \alpha y + k\sigma(xy - x^2). \qquad (3/6)$$

From a study of the submission of Germany in the 1920s I estimated that :

$$\sigma > 0{\cdot}34 \times 10^{-6} \text{ per person}, \qquad (3/7)$$

(Richardson (23), 1949). As this information is meagre, let us change to new variables which involve σ in such a way that we can discuss its general effects without knowing its numerical value. Thus define U, V by

$$x = U/\sigma, \; y = V/\sigma, \qquad (3/8), (3/9)$$

so that U and V are pure numbers. Eqn. (3/5) and (3/6) then become

$$dU/dt = -\alpha U + kV + k(UV - V^2) \qquad (3/10)$$

$$dV/dt = kU - \alpha V + k(UV - U^2). \qquad (3/11)$$

Further let

$$t = \tau/k, \qquad (3/12)$$

so that τ is a pure number. For thus

$$\frac{dU}{d\tau} = -\frac{\alpha}{k}U + V + UV - V^2, \qquad (3/13)$$

$$\frac{dV}{d\tau} = U - \frac{\alpha}{k}V + UV - U^2. \qquad (3/14)$$

Only one constant remains explicit, namely, α/k, and it is a pure number. Before 1877 arms-races did not occur, and statesmen believed in a stable balance of power, such as would occur if $\alpha/k \geqslant 1$. From a study of the arms-race which led to the First World War it was concluded that $k - \alpha =$

$0{\cdot}2_7$ year^{-1}, whereas the disarmament after it gave $\alpha = 1{\cdot}1_3$ years^{-1} on the average (Richardson, 1949). Whence $\alpha/k = 0{\cdot}8$. The disarmament after the Second World War gave $\alpha = 1{\cdot}9$ years^{-1}. The present arms-race, with considerably more uncertainty, gives $k - \alpha = 2{\cdot}2$ years^{-1}. These last two, taken together, give $\alpha/k = 0{\cdot}46$ (Richardson (26), 1953). There is thus some evidence that α/k, has decreased, during the present century. The formulæ for any value of α/k are stated above; but in diagrams and in numerical examples the modern situation will be illustrated by

$$\alpha/k = \frac{1}{2}. \qquad (3/15)$$

Arms-races between equal nations show most conspicuously in the sum of the warlike efforts of the two sides, whereas submission is essentially concerned with their difference. Let us therefore define S and D by

$$S = U + V, \quad D = V - U. \qquad (3/16), (3/17)$$

By addition and subtraction of eqn. (3/13) and (3/14) it follows that

$$\frac{dS}{d\tau} = S\left(1 - \frac{\alpha}{k}\right) - D^2, \quad \frac{dD}{d\tau} = -D\left(1 + \frac{\alpha}{k}\right) + DS, \qquad (3/18), (3/19)$$

or for a modern illustration

$$dS/d\tau = \frac{1}{2}S - D^2, \quad dD/d\tau = -\frac{3}{2}D + DS. \qquad (3/20), (3/21)$$

We can still keep track of submissiveness by remembering that it comes in by way of the quadratic terms, and would cancel in the linear term.

These equations purport to represent a wide range of possible behaviour. In particular, suppose that initially the nations were considerably disarmed, as in 1948, or more so. The point $S = 0$, $D = 0$ is a point of balance. Sufficiently close to it the quadratic terms are negligible, so that

$$S = \text{const}_1\, e^{\frac{1}{2}\tau}, \quad D = \text{const}_2\, e^{-\frac{3}{2}\tau}. \qquad (3/22), (3/23)$$

The exponential increase of S is the arms-race beginning inconspicuously. Meanwhile D tends to zero, so that the opposing sides tend to become equal in their warlike efforts. But (3/21) may be written

$$\frac{d \log |D|}{d\tau} = S - \frac{3}{2}. \qquad (3/24)$$

As S grows, $\log|D|$ would at first become more negative; but there would come a time τ_1, say, when $S = 3/2$, and immediately after that D would be unstable. That is how submissiveness would first begin to reveal its presence.

Too much determinism can be avoided by supposing that D had become near to zero; so that a chance disturbance could change its sign before or soon after it became unstable. After D^2 had grown sufficiently, $\frac{1}{2}S - D^2$ would become negative and so S would decrease. The decrease of S would be recognized as the end of the arms-race. The above argument was in effect that of my letter to *Nature* (25) of September 29th, 1951; and I still think it credible in the restricted part of the S, D plane to which it refers.

But now comes the difficulty. Dr. Cartwright, by sketching the integrals over a wider field, emphasized a theoretical point of balance from which the trajectories spiralled outwards. No corresponding historical facts have been found. There is the further difficulty that if the representative particle always comes out, how did it ever get in? There is a point of balance if simultaneously

$$0 = \frac{1}{2}S - D^2, \quad 0 = -\frac{3}{2}D + DS. \qquad (3/25), (3/26)$$

The solution at $S = 0$, $D = 0$ was well known. It represented general disarmament, but was an unstable 'in-and-out' point. Other solutions are:

$$S = \frac{3}{2}, \quad D = \pm\frac{1}{2}\sqrt{3} = \pm 0{\cdot}866. \qquad (3/27), (3/28)$$

That is either (U, V) or $(V, U) = 1{\cdot}183, 0{\cdot}317$.

The general features of the theoretical field are shown in Fig. 1, which was obtained by the method used by B. van der Pol in solving his relaxation equation ((18), 1926). The lines, called 'isoclines,' along each of which dD/dS is constant, were drawn first. The isoclines are all conic sections, and each of them goes through both points of balance. For example, one of these conics is a pair of intersecting straight lines, each line passing through one point of balance. Points on each isocline were calculated and plotted, and a smooth curve was drawn through them. In contrast with that fairly definite construction, the drawing of the trajectories was in effect an integration, for it was a fitting together of pieces to form a continuous curve; and this process involved more uncertainty. However, van der Pol ((18), 1926) used it with good effect.

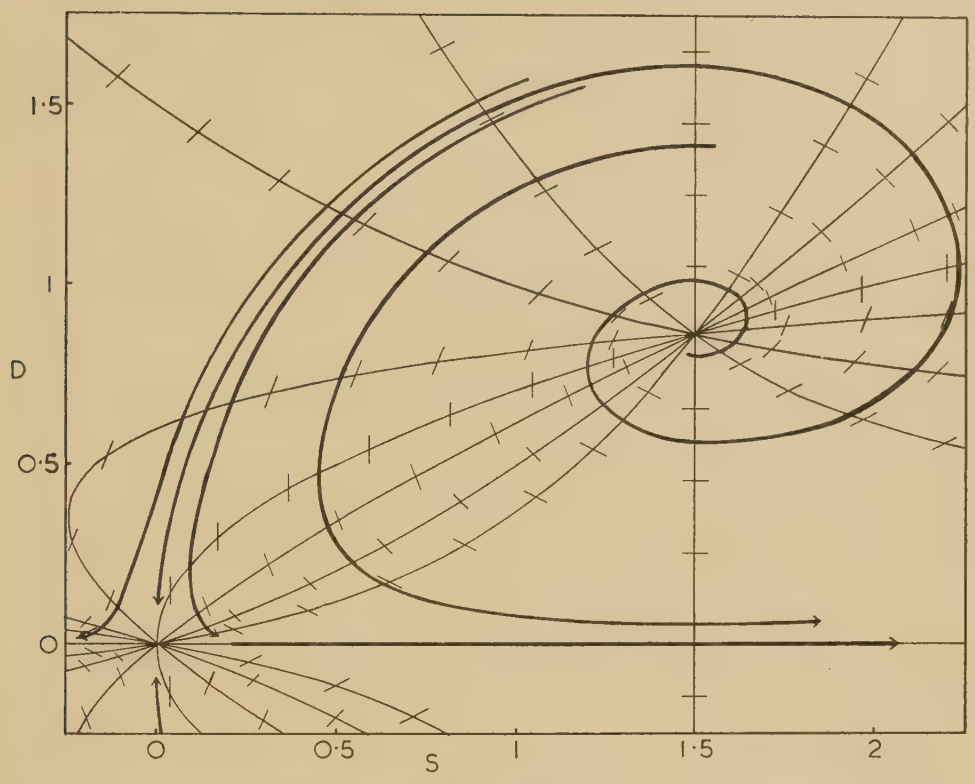

FIG. 1.—On the first hypothesis this shows the field of two opposing sides that are equal in their constants. S is proportional to the sum and D to the difference of their warfinpersals. The thin curves are the isoclines along each of which dD/dS is constant. They are drawn for $dD/dS = 0, \frac{1}{2}, 1, \frac{7}{2}, \infty, -\frac{7}{2}, -1, -\frac{1}{2}$. The directions of motion are indicated by short segments crossing each isocline. The thick lines are trajectories drawn so as to fit the prescribed directions.

To investigate in more detail the neighbourhood of the point of balance at $S = \frac{3}{2}$, $D = \frac{\sqrt{3}}{2}$, define ε and ζ by

$$S = \frac{3}{2} + \varepsilon, \quad D = \frac{\sqrt{3}}{2} + \zeta, \qquad (3/29), (3/30)$$

and substitute the expressions in the eqn. (3/20) and (3/21) obtaining

$$\frac{d\varepsilon}{d\tau} = \frac{1}{2}\varepsilon - \sqrt{3}\zeta - \zeta^2, \quad \frac{d\zeta}{d\tau} = \frac{\sqrt{3}}{2}\varepsilon + \varepsilon\zeta. \qquad (3/31), (3/32)$$

The terms of zero degree have cancelled. On the circle $\varepsilon^2 + \zeta^2 = 1$ the terms of the second degree are of the same order of magnitude as those of the first degree. Much within that, say, inside the circle, $\varepsilon^2 + \zeta^2 = (1/10)^2$, the terms of the second degree become almost negligible. This circle of radius 1/10, to the interior of which the following approximation is restricted, is quite a small part of Fig. 1. In the standard manner try the assumption that, where A, B are constants,

$$\varepsilon = Ae^{\lambda\tau}, \quad \zeta = Be^{\lambda\tau}, \qquad (3/33), (3/34)$$

then
$$A\lambda = \frac{1}{2}A - \sqrt{3}\,B, \qquad (3/35)$$

and
$$B\lambda = \frac{\sqrt{3}}{2}A. \qquad (3/36)$$

These simultaneous equations for A and B are consistent if the determinant

$$\begin{vmatrix} \frac{1}{2}-\lambda & , & -\sqrt{3} \\ \frac{\sqrt{3}}{2} & , & -\lambda \end{vmatrix} = 0 ; \qquad (3/37)$$

that is, if

$$\lambda^2 - \frac{1}{2}\lambda + \frac{3}{2} = 0, \qquad (3/38)$$

the roots of which are

$$\lambda_1, \lambda_2 = 0{\cdot}25 \pm i\, 1{\cdot}199. \qquad (3{\cdot}39)$$

Because these are complex with positive real part therefore the particle moves spirally outwards. The period of ε or ζ is $2\pi/1{\cdot}199$ on the scale of τ. But $t = \tau/k$; and k has been found from the data of 1949, 1950, 1951 very doubtfully to be $4{\cdot}_1$ years^{-1}. Therefore the period is :

$$\frac{2\pi}{1{\cdot}199 \times 4{\cdot}1} \text{ years} = 1{\cdot}_3 \text{ years}. \qquad (3/40)$$

For a political double ' swing of the pendulum ' that seems incredibly rapid. Maybe the warfinpersals for 1952 and 1953, when available, will show that k is less and the period greater. A method for making an accurate drawing of the trajectories in the neighbourhood of any point of balance is worked out in the author's *Arms and Insecurity* ((23), 1949, § 3/2/4). When the roots λ are complex, the trajectory can be obtained from an auxiliary equiangular spiral by stretching or squeezing it at right angles to a characteristic direction E. Application of those general formulæ to the present case shows that the radius vector of the equiangular spiral revolves counter-clockwise and that its length increases in the ratio 3·707 per turn, or in the ratio 1·388 per quarter turn. The characteristic direction is at $E = 105°$ from the ε-axis. The trajectory is obtained by stretching the equiangular spiral in the ratio 1·5005.

Dr. Chike Obi (17), of University College, Ibadan, Nigeria, who was formerly a pupil of Dr. Cartwright, has made an extensive study, from the point of view of pure mathematics, of the equation

$$\ddot{x} + (\alpha + \beta x)\,\dot{x} + x + \gamma x^2 = 0, \qquad (3/41)$$

where α, β, γ are constants not too far removed from zero. It is possible by a chain of transformations to connect the present hypothesis exactly to Obi's form. For this purpose we first introduce a new variable $\omega = D^2$, then shift the origin to the vortex point, and finally eliminate that deviation which is not in the S direction.

It is sometimes asserted that ' history repeats itself.' If so, we should see a closed orbit. Actually there are only spirals in Fig. 1.

The origin is what I have elsewhere (1949) named an ' in-and-out ' point. Lefschetz ((12), 1948) names it a ' col.' Cartwright calls it a ' saddle point ' or ' col.' To the right of it the drift is towards rearmament ; to the left of it the drift is towards negative armament, which I formerly suggested ((21), 1939, p. 7) was another name for co-operation. The practical difficulty was : how could the nations ever get across to the co-operative side, seeing that the flow diagram does not permit a direct crossing ? Fig. 1, however, shows that by going round over the top of the vortex there is a regular way into the co-operative region. These are strange ideas. They all purport to refer to times when there was no fighting.

There is a boundary-spiral which runs right into the ' in-and-out ' point. Any slightly lesser spiral passes first through the arms-race. Any slightly greater spiral goes beyond the ' in-and-out ' point into the co-operative region.

This vortex, which certainly follows from hypothesis I, is very difficult to accept as a description of what nations in fact did. Herein lies a good example of the usefulness of mathematics in social science. The consequences of the hypothesis were not all immediately obvious. In the familiar region they seemed plausible. It was only by deducing consequences in an unfamiliar region that incredible entailments were noticed.

Let us therefore pass on to a different hypothesis.

IV. SECOND HYPOTHESIS ABOUT SUBMISSIVENESS

As in the first hypothesis, the defence-coefficient must become negative when the threat becomes overwhelming ; but this change of sign will now be contrived differently. One of the reasons given ((21) 1939, p. 23) for the theory which has just been abandoned was that " the

overwhelmingness of the threat depends not on y absolutely, but on y relative to x." The reconsideration is that it is not relative to the x present at that time, but to the x that could be mobilized. This 'x that could be mobilized' is almost a constant for any one nation, but varies very much from one nation to another.

Thus let the new hypothesis be

$$dx/dt = g - \alpha x + ky\{1 - y/m\}, \qquad (4/1)$$

$$dy/dt = h - \beta y + lx\{1 - x/n\}, \qquad (4/2)$$

where m and n are constants for those nations, and are such that m is proportional to the total warfinpersal that the x-nation could mobilize, and n is proportional to the corresponding quantity for the y-nation. Therefore

$$x \leqslant m, \qquad y \leqslant n. \qquad (4/3), (4/4)$$

As before, g and h will be neglected, because they are not essential to the phenomena of submissiveness.

Consequences of the Second Hypothesis (a) *For Very Unequal Nations.*—Suppose that $m \geqslant n$, then n/m, and *a fortiori* y/m, is near zero; so that $1 - y/m$ is almost unity. There is, on the contrary, no reason why x/n should not much exceed unity. In these circumstances we have approximately from (4/1) and (4/2)

$$dx/dt = -\alpha x + ky, \qquad (4/5)$$

$$dy/dt = -\beta y - lx^2/n. \qquad (4/6)$$

Therefore y is decreasing. The smaller nation is suppressed. For this case of very unequal nations there is nothing much to choose between the two hypotheses. Either of them describes what usually happens in a world of power politics.

(b) *For Nations Equal in their Constants.*—From (4/1) and (4/2) we have

$$dx/dt = -\alpha x + ky\,(1 - y/m), \qquad (4/7)$$

$$dy/dt = -\alpha y + kx\,(1 - x/m). \qquad (4/8)$$

The constant m can be made to disappear by a change of variables. Thus define u, v by

$$x = um, \qquad y = vm; \qquad (4/9), (4/10)$$

so that

$$u \leqslant 1, \qquad v \leqslant 1. \qquad (4/11), (4/12)$$

For then

$$du/dt = -\alpha u + kv - kv^2, \qquad (4/13)$$

$$dv/dt = -\alpha v + ku - ku^2. \qquad (4/14)$$

The two remaining constants, α and k, can be reduced to one, in the following way:

As in the former theory, define τ by

$$t = \tau/k; \qquad (4/15)$$

so that

$$\frac{du}{d\tau} = -\frac{\alpha}{k}u + v - v^2, \qquad (4/16)$$

$$\frac{dv}{d\tau} = -\frac{\alpha}{k}v + u - u^2. \qquad (4/17)$$

Finally, define S and D by

$$S = u + v, \qquad D = v - u, \qquad (4/18), (4/19)$$

obtaining

$$\frac{dS}{d\tau} = \left(1 - \frac{\alpha}{k}\right)S - \frac{1}{2}(S^2 + D^2), \qquad (4/20)$$

$$\frac{dD}{d\tau} = -\left(1 + \frac{\alpha}{k}\right)D + SD. \qquad (4/21)$$

Or, with $\alpha/k = \frac{1}{2}$ as in the former theory, to represent modern conditions

$$dS/d\tau = \frac{1}{2}S - \frac{1}{2}(S^2 + D^2), \qquad (4/22)$$

$$dD/d\tau = -\frac{3}{2}D + SD. \qquad (4/23)$$

The equation for $dD/d\tau$ is exactly the same as on the former hypothesis; but that for $dS/d\tau$ has $\frac{1}{2}(S^2 + D^2)$ in place of simply D^2. The points of balance where $dS/d\tau$ and $dD/d\tau$ vanish simultaneously are given by

$$0 = S - S^2 - D^2, \qquad 0 = D\left(S - \frac{3}{2}\right). \qquad (4/24), (4/25)$$

If from (4/25) we accept $D = 0$ and substitute in (4/24) we get $0 = S - S^2$ which has solutions $S = 0$ or $S = 1$. Alternatively, if $S = 3/2$, the second condition is satisfied and the first becomes $D^2 = -3/4$, so that D is imaginary. There are therefore only two points of balance, namely,

$$S = 0,\ D = 0 \text{ and } S = 1,\ D = 0. \qquad (4/26), (4/27)$$

We seem well rid of the puzzlingly unsymmetrical point of balance of the former theory. The point $S = 0$, $D = 0$ is of course of the same character as before, namely, an in-and-out point.

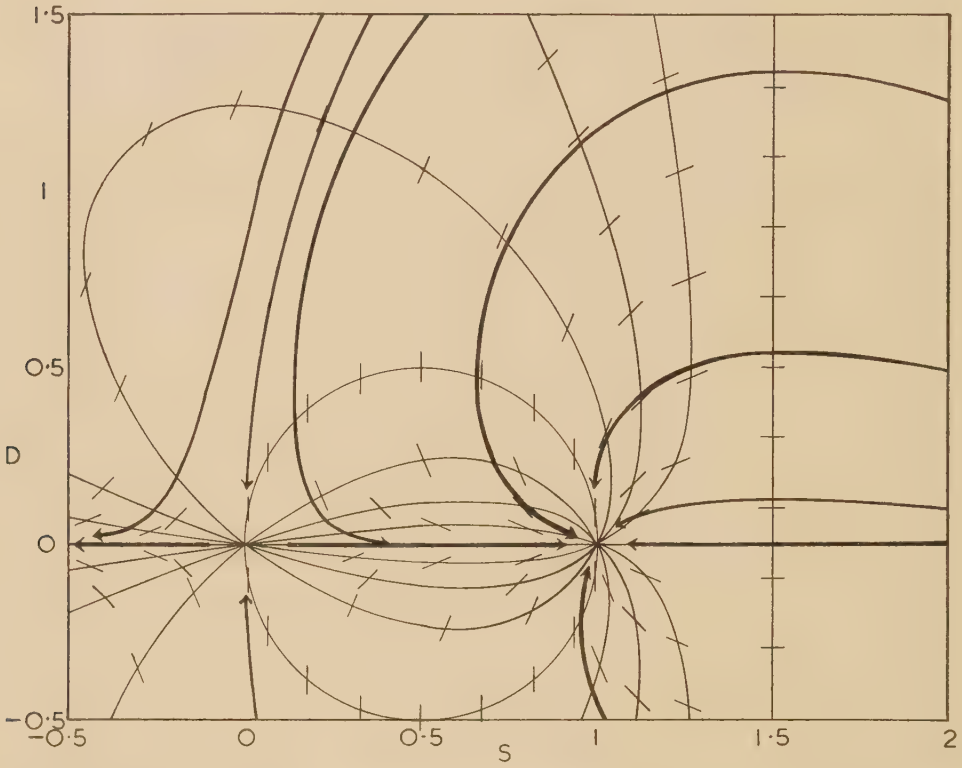

FIG. 2.—The field of two opposing sides on the second hypothesis. As in Fig. 1, S is proportional to the sum and D to the difference of their warfinpersals. The sides are equal as to their constants. The thin curves are the isoclines, along each of which dD/dS is constant. The direction of motion is indicated by a short segment. The thick curves are trajectories, along any one of which the point which represents the two warfinpersals could move.

The general field is shown in Fig. 2 by way of isoclines and trajectories. The isocline

$$\frac{dD}{dS} = j\,, \qquad (4/28)$$

where j is any constant, works out to

$$S - S^2 - D^2 = \frac{D}{j}(2S - 3)\,; \qquad (4/29)$$

so that changing the sign of j has the same effect as changing the sign of D.

In the region of the arms-race, where D tends to zero while S increases, we may neglect D in (4/22) which then simplifies to

$$\frac{dS}{d\tau} = \frac{1}{2}S - \frac{1}{2}S^2\,, \qquad (4/30)$$

in which the variables are separable. Its integra lcan be verified to be

$$\frac{1}{S} = 1 + Ce^{-\tau/2}, \qquad (4/31)$$

where C is an arbitrary constant. So as $\tau \to \infty$, $S \to 1$. Let us investigate the neighbourhood of the point $S = 1$, $D = 0$ on the second hypothesis. Define ε here differently from the ε in (3/29), and thus

$$S = 1 + \varepsilon. \qquad (4/32)$$

Insert this in (4/22), (4/23). Then

$$d\varepsilon/d\tau = -\frac{1}{2}\varepsilon - \frac{1}{2}(\varepsilon^2 + D^2), \qquad (4/33)$$

$$dD/d\tau = -\frac{1}{2}D + \varepsilon D. \qquad (4/34)$$

If we neglect the quadratic terms in ε and D, the solutions are

$$\varepsilon = Ae^{-\tau/2}, \qquad D = Be^{-\tau/2}, \qquad (4/35), (4/36)$$

where A, B are arbitrary constants. This point is therefore one of thoroughly stable balance.

The theoretical state of affairs at which we have arrived as a deduction from the second hypothesis might please some people : large armed forces, stable balance, and therefore no fighting. One of the editors of the *Cambridge Modern History* (3), Stanley Leathes, wrote in his preface to Vol. XII : " On the whole, the existence of this tremendous military equipment makes for peace. The consequences of war would be felt in every household ; and statesmen, as well as nations, shrink from the thought of a conflict so immense." This was published in 1910. Unfortunately for Leathes' theory, the First World War broke out four years later. Yet this theory persists. Winston Churchill, speaking at the Pilgrims' dinner on October 14th, 1952, said that " a third world war is unlikely because both sides know that it would begin with horrors never dreamt of before."

So my second hypothesis about submissiveness appears to agree with Leathes and with Churchill, but to have the outbreak of the First World War as evidence against it. Personally I have no confidence in it, regarding it as wishful thinking.

V. THIRD HYPOTHESIS (FOR SIDES EQUAL IN THEIR CONSTANTS)

When formulating the two former hypotheses, I had hoped to have some insight into what the motives might be. Now I ignore motives, and try to make a mathematical description of the courses in time of the opposing warfinpersals before, during, and after the First World War. The facts must first be reviewed.

(i) From A.D. 1907 to 1914, the two sides being the Triple Entente and the Triple Alliance, the representative point was moving nearly in a straight line with a speed proportional to its distance from a " point of balance " at $x = £135$ million, $y = £73$ million. The graph is given by Richardson ((26), 1953). This is called an arms-race. The sterling has been converted to warfinpersal by division by £97·3 per year ; that being a British wage the conversion is very rough for the other nations. Before these numbers were plotted in Fig. 3, x and y were interchanged.

(ii) There followed the First World War. Italy changed sides. The warfinpersal of Britain attained a maximum of 12·2 million equivalent persons in 1917 (Richardson (26), 1953). I have not succeeded in directly ascertaining the warfinpersals of the other belligerents, but a rough indirect estimate can be made as follows. Lloyd George ((13), *War Memoirs*, 1938, p. 1575) states the composition of the British Army in March, 1918 (excluding Colonial and Indian troops). From this statement I compute that there were 2·96 million British combatants. Now 12·2/2·96 = 4·13. It is understandable that the warfinpersal might considerably exceed the combatant strength, because the warfinpersal accounts for the non-combatants and the munition makers. Lloyd George (p. 1,577) further states that at the end of 1917 the total combatant strength of the Allies, without including any Russians or Roumanians, was 5·4 million against 5·2 million for the Central Powers. My rough estimate is made by assuming that the British factor, 4·13, applied also to the other nations concerned, so that the maximum warfinpersals were : Allies $22{\cdot}3 \times 10^6$, Central Powers $21{\cdot}5 \times 10^6$, total $43{\cdot}8 \times 10^6$ equivalent persons.

Because Germany ultimately submitted, it is relevant to enquire when the first hint of submissiveness appeared in the German state as a whole. The question is not about individuals or

groups. In another place ((22, 1948) I have given a theory of war-moods in which the people in German groups are represented by five variables ; but here only one variable is available. On December 12th, 1916, the German Chancellor suggested a peace conference. The opposing Powers, after consultation, said they " refuse to consider a proposal which is empty and insincere." (Lloyd George, p. 661.) There were various other tentative and secret peace feelers in 1917 and 1918, but mostly Austrian. According to a Reichstag Committee, quoted by Lloyd George (p. 1932), " Up to 15th July, 1918, the Supreme Army Command rejected the view that victory was no longer possible of attainment by force of arms, and gave no support to peace negotiations upon the basis of a military stalemate . . ." During the next three weeks the fresh Americans came to the aid of the tired French and British. A battle on August 8th, 1918, in which the Germans had no reserves left, changed Ludendorff's opinion, and on August 14th the Kaiser, presiding over a Crown Council at Spa, admitted that Germany would have to find a suitable moment in which to come to an understanding with the enemy.

(iii) After the First World War, the sides were not quite the same. There was the Russian revolution and its consequences. U.S.A. withdrew into isolation. A new diagram was drawn showing the warfinpersal of Germany plotted against that of its entourage, consisting of France, Britain, Poland, Czechoslovakia, and Russia. This appears in a microfilm (Richardson (23), 1949) and is modified here into part of Fig. 3. It shows motion, along a nearly straight track and slowing down as a point of balance is approached. There was a delay of eleven years in the neighbourhood of this point, and then a departure, with increasing speed along another almost straight track. This was the second great arms-race beginning.

(iv) It led to the Second World War during which the warfinpersals approached the state of total mobilization, but were difficult to ascertain. The sides were rather different from those of the first war, Japan and Italy in particular having changed over previously, and Italy changing back again. In October, 1939, Hitler made a peace offer ; but Britain and France did not regard it as at all submissive. After that the demand for the unconditional surrender of Germany delayed any sign of submissiveness until defeat.

(v) The Occupation Statute for Germany, an expression of submission, has now been in operation for seven years. Disarmament after the Second World War was barely over when enmity between U.S.A. and the Soviet Union began to dominate the scene.

Fitting Formulæ to the Facts.—It seems permissible to ignore the changes in the membership of the sides, so that x and y may be regarded as having the same meaning throughout. This has already been assumed in Fig. 3.

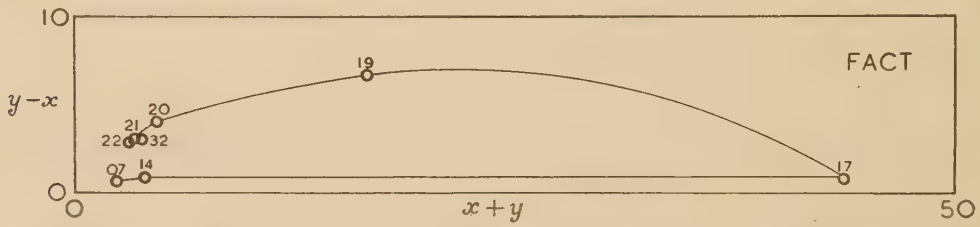

FIG. 3.—Pro- and anti-German warfinpersals in Europe, 1907 to 1932, simplified by ignoring other changes in the sides ; and in particular by ignoring the brief, but balance-altering, intervention of U.S.A. The co-ordinates are the sum and difference. The numbers set beside the plotted points are the dates.

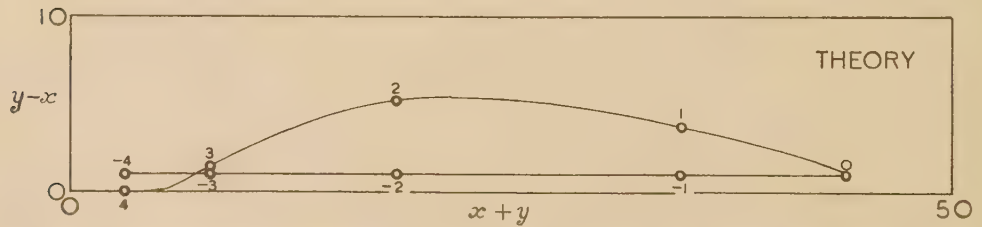

FIG. 4.—A mathematical representation, plotted from formulæ (6/17), (6/19), and (6/20) of the warfinpersals before, during, and after the First World War. For comparison with Fig. 3. Numbers set beside the plotted points are values of t. A unit of one year makes Fig. 4 agree roughly with Fig. 3 as to time.

As the mathematical description is to be of that which usually happened, we are led to beg the question "Could an arms-race end without fighting?" by assuming, on the meagre evidence of two out of three trials, that it usually does not. This assumption is in strong contrast with the first and second hypotheses, which leave that important question open at the outset.

Remembering that an arms-race has to do mainly with the sum $x + y$ and submission mainly with the difference $y - x$ of the warfinpersals, let ξ and η be defined thus:

$$\xi = x + y, \qquad \eta = y - x, \tag{6/1), (6/2}$$

so that conversely

$$x = \frac{1}{2}(\xi - \eta), \qquad y = \frac{1}{2}(\xi + \eta). \tag{6/3), (6/4}$$

Accordingly ξ and η are respectively constant multiples of the S and D of former theories. We can now describe the changes in ξ without mentioning η, thus:

(i) After wars there have been disarmaments. Therefore $d\xi/dt$ must be capable of either sign.

(ii) For small ξ it is observed that $d\xi/dt$ varied as $\pm (\xi - \xi_a)$ where ξ_a is a constant, called a point of balance.

(iii) There was an upper bound to ξ when both the opposing sides were fully mobilized for war. Let γ denote this upper bound.

The above three requirements are satisfied by the assumption that

$$\left(\frac{d\xi}{dt}\right)^2 = a^2(\xi - \xi_a)^2(\gamma - \xi), \tag{6/5}$$

where a^2 is some constant. A simpler expression could not be found; η does not appear; ξ cannot exceed γ, for if it did then $d\xi/dt$ would become imaginary; when ξ is near zero then $d\xi/dt \doteq \pm a(\xi - \xi_a)\sqrt{\gamma}$.

The variables are separable, so that (6/5) can be integrated. The result, which can be verified by differentiation, is

$$\xi - \xi_a = (\gamma - \xi_a)\,\mathrm{sech}^2\, Bt, \tag{6/6}$$

where

$$B = \frac{a\sqrt{(\gamma - \xi_a)}}{2}. \tag{6/7}$$

At $t = 0$, $\xi = \gamma$; that is to say, the constant of integration has been chosen so that $t = 0$ at the maximum of ξ, which occurred during the war. As $t \to -\infty$, $\xi \to \xi_a$: there is not, on this theory, any sudden or definite beginning of the increase of warfinpersal. This vaguely agrees with the author's general impression of a collection of deadly quarrels that ended since A.D. 1820 (Richardson (24), 1950). The question "Who began it?" seldom had any non-controversial answer. He abandoned any attempt in that collection to name the aggressors.

To determine B from observation, let t' be the time at which ξ was halfway between γ and ξ_a. It follows from (6/6) that

$$B = \frac{0{\cdot}8814}{t'}. \tag{6/8}$$

The theoretical diagram, Fig. 4, has been drawn with

$$B = 0{\cdot}5 \text{ year}^{-1}. \tag{6/9}$$

Next as to η, the facts represented in Fig. 3 suggest that η did not alter much while ξ was increasing, but that, after the maximum of ξ, changes in η were conspicuous. These were of two kinds: at first, while ξ was large, $|\eta|$ increased; later when ξ was near ξ_b there was a settling of η to about the same value ξ_b, so that finally

$$x = \frac{1}{2}(\xi - \eta) = 0, \tag{6/10}$$

$$y = \frac{1}{2}(\xi + \eta) = \xi_b. \tag{6/11}$$

This expresses the final submission of the x-side. It would have been a mistake to have made η tend to zero, unless ξ also did. To express both kinds of change in η we need something like a disjunction or "either-or" formula. One can be concocted on the model of $\xi + 1/\xi$ which behaves like ξ when ξ is large, but like $1/\xi$ when ξ is near zero. When ξ was large, η departed further from zero without changing sign, as if $d\eta/dt$ were proportional to η. When ξ was small, η tended to ξ_b, as if $d\eta/dt$ were proportional to $\xi_b - \eta$. These diverse tendencies could be combined by weighting them respectively by ξ and $1/\xi$. That, however, would make the two terms of different

dimensions, for ξ is a number of equivalent persons. So also is γ. It is found to be satisfactory to use the dimensionless weights $2\xi/\gamma$ and $\gamma/2\xi$. Thus one arrives at the assumption that

$$\frac{d\eta}{dt} = \sigma_3 \left\{\eta\left(\frac{2\xi}{\gamma}\right) + (\xi_b - \eta)\left(\frac{\gamma}{2\xi}\right)\right\}, \qquad (6/12)$$

in which σ_3 is some constant measure of submissiveness, and is the reciprocal of a time.

A most important fact about the submissiveness of Germany as a whole was that it came on suddenly. Sudden commencements have been much studied mathematically in connexion with seismographs, and with the switching on of electric circuits. Heaviside introduced a function $H(t)$ defined to be zero for $t < 0$, but unity for $t > 0$ (H. and B. S. Jeffreys (30), 1946, p. 219). If the onset is required to be not quite so sudden, four approximations to $H(t)$ are available (Van der Pol and Bremmer (19), 1950, pp. 56–7). I shall write

$$\frac{d\eta}{dt} = H(t)\,\sigma_3 \left\{\eta\,\frac{2\xi}{\gamma} + (\xi_b - \eta)\,\frac{\gamma}{2\xi}\right\}. \qquad (6/13)$$

So that $d\eta/dt = 0$ for $t < 0$.

The inevitable non-linearity has been introduced, but in a manner quite different from that of hypotheses I and II.

As a discontinuity at $t = 0$ in $d\eta/dt$ has been admitted, a discontinuity also in $d^2\xi/dt^2$ does not seem too fanciful. It allows us to joint together at $t = 0$ two formulæ for ξ, both like (6/6), but the later one having b for a and C for B, so that the final point of balance ξ_b need not agree with the initial ξ_a. Observed constants, in million equivalent persons, were about

$$\gamma = 44, \qquad \xi_b = 3. \qquad (6/14), (6/15)$$

The integration of (6/13) is, however, much simplified if we take $\xi_b = 0$, for then η is a factor of the second member, so that for $t > 0$

$$\frac{d\log|\eta|}{dt} = \sigma_3 \left\{\frac{2\xi}{\gamma} - \frac{\gamma}{2\xi}\right\}. \qquad (6/16)$$

This simplified problem is solved as a preliminary. Substitute, in view of (6/6) and (6/9),

$$\xi = \gamma \operatorname{sech}^2 \frac{t}{2}, \qquad (6/17)$$

obtaining from (6/16)

$$\frac{d\log|\eta|}{dt} = \sigma_3 \left\{2\operatorname{sech}^2 \frac{t}{2} - \frac{1}{2}\cosh^2 \frac{t}{2}\right\}. \qquad (6/18)$$

The integral is

$$\frac{\eta}{\eta_0} = \operatorname{Exp}\left[\sigma_3 \left\{4\tanh\frac{t}{2} - \frac{1}{4}\sinh t - \frac{1}{4}t\right\}\right], \qquad (6/19)$$

which can be verified by differentiation. The measure σ_3 of submissiveness was, of course, unknown ; but Fig. 4 was plotted with

$$\sigma_3 = 1 \qquad (6/20)$$

and comparison of Fig. 4 with Fig. 3 shows that $\sigma_3 = 1$ is of the right order of magnitude. The complete eqn. (6/13), with ξ_b not zero, has for an integrating factor the expression (6/19) (Forsyth (7), 1929, p. 20) and is thus reducible to quadratures. But these look difficult and are perhaps not worth pursuing at this stage.

VI. CONCLUSION

The memory of recent overwhelming military defeat, say, within five years, certainly leaves a tendency to submit. A large state has been known to overawe a very much smaller state, without any fighting. Apart from those two types, I have looked for submissive effects, but have not noticed any. Two hypotheses as to how submissive effects might arise between rested and approximately equal states, called A and B, are considered. Mathematical hypotheses like these are too complicated to be expressed, both briefly and accurately, in words. However, one can say that the description of an arms-race, as caused by mutual stimulation, is modified by supposing either (i) that the overwhelming nature of the threat

against A is proportional to the excess of B's armed force over A's at the same time, or (ii) that it is not A's armed force at the same time that matters, but what A could mobilize. Not all the consequences of these hypotheses are immediately obvious. In historically known circumstances the hypotheses seemed plausible. It was only by deducing their consequences in unusual circumstances that incredible entailments were noticed. This is a good example of the usefulness of mathematics in social science.

Most governments seem to expect that a show of armed force, to a rested and equal group of nations, will have an effect of the same sign as an overwhelming military defeat. This expectation appears to be a mistake : the actual effects go opposite ways.

The Quakers, Gandhi, and others have tried to persuade everybody to be as non-violent as the exceptional Christian saints. Most people feel the suggestion to be too utterly contrary to their defensive impulses. Could not a middle way be found, with the motto " defence without provocation." This motto might well be applied to words as well as deeds. The existence of the various arms of nation A presumably have different psychological effects in peacetime on nation B ; for example, one may guess that B is much more provoked to retaliate by the existence of A's long-range bombers than by A's land-based fighters. Jonathan Griffin (8) discussed ' defence without menace ' in 1936, and now again the problem of the classification of armaments into provocative and non-provocative requires much more attention than it receives.

Submission, except in the sense of defeat, is not a spectacular subject for newspaper headlines ; it is not like fighting for freedom. Yet we need to remember that in any organization, whether it be a social club, a scientific society, a business, a national parliament, or the United Nations, a suitable amount and distribution of submissiveness is essential for smooth working. If each member struggled all the time for his own advantage there would be chaos.

The present study has perhaps over-emphasized the bitter submission of the defeated, threatened, suppressed, or snubbed, and has neglected the joyous devotion to an admired leader. To correct the balance let me instance a classical story illustrating the latter tendency, that of Naomi and Ruth. It is described as an affair of tribal loyalties.

In the United Nations, activities of the follow-my-leader type are similarly noticeable on both sides of the major schism.

REFERENCES

1. Allport, G. W. (1945). *Personality.* New York : Henry Holt.
2. Bateson, G. (1936). *Naven.* Cambridge : University Press.
3. *Cambridge Modern History*, XII (1910). Cambridge : University Press.
4. Cartwright, M. L. (1949). Report on Non-Linear Vibrations. *The Advancement of Science*, **6**.
5. Cartwright, M. L. (1952). *Non-Linear Vibrations. Math. Gazette*, **36**, 81–8.
6. Cartwright, M. L. (*no date*). *Forced oscillations in non-linear systems.* (Lectures at Professor Lefschetz's seminar.)
7. Forsyth, A. R. (1929). *A Treatise on Differential Equations.* London : Macmillan.
8. Griffin, J. (1936). *Alternative to Rearmament.* London : Macmillan.
9. Horne, M. R. (1951). Letter in *Nature*, November 24th.
10. *Keesing's Contemporary Archives.* Bristol : Keesing's Publications : cited as ' K.'
11. Landau, H. G. (1951). *Bull. Math. Biophysics*, 13, 1–19.
12. Lefschetz, S. (1948). *Lectures on Differential Equations.* Princeton : University Press.
13. Lloyd George, D. (1938). *War Memoirs.* London : Odhams.
14. Lorenz, K. Z. (1952). *King Solomon's Ring.* London : Methuen.
15. McDougall, W. (1917). *Social Psychology.* London : Methuen.
16. Moyal, J. E., Bartlett, M. S., and Kendall, D. G. (1949). ' Symposium on Stochastic Processes.' *J. Roy. Stat. Soc.*, B., II, 150–282.
17. Obi, C. (1950). (Unpublished.) Thesis for Ph.D. : Cambridge University.
18. Pol, B. van der (1926). *Phil. Mag.*, II, 978–92.

19. Pol, B. van der, and Bremmer, H. (1950). *Operational Calculus based on the Two-sided Laplace Integral.* Cambridge: University Press.
20. Richardson, L. F. (1935). *Nature*, **135**, 830.
21. Richardson, L. F. (1939). ' Generalized Foreign Politics.' *Brit. J. Psychol. Mon. Sup.*, No. 23.
22. Richardson, L. F. (1948). *Psychometrika*, **13**, 147–74 and 197–232.
23. Richardson, L. F. (1949). *Arms and Insecurity* (microfilm, published by the author).
24. Richardson, L. F. (1950). *Statistics of Deadly Quarrels* (microfilm, published by the author).
25. Richardson, L. F. (1951). *Nature*, **168**, 567.
26. Richardson, L. F. (1953). *The Indian Journal of Statistics*, **12**, 205–28.
27. Richardson, L. F., and Stommel, H. (1948). *Journ. Meteor.*, 5, 238–40.
28. Russell, Bertrand (1914). *War the Offspring of Fear.* London: Union of Democratic Control.
29. Volterra, V. (1930). *Theory of Functionals.* London: Blackie.
30. Jeffreys, H. and B. S. (1946). *Methods of Mathematical Physics.* Cambridge: University Press.
31. Weierstrass, K. (1875). *K. Akad. der Wiss.* (October 28th). Also *Werke*, II, 75–6.

Vol. VI
Part II

The British Journal of Statistical Psychology

November,
1953

A PRELIMINARY STUDY OF THE THEMATIC APPERCEPTION TEST

By AMYA SEN
University College, London

I. *History of Method.* II. *Subjects.* III. *Test Material and Mode of Scoring.* IV. *Reliability and Validity of Assessments.* V. *Summary and Conclusions.*

I. HISTORY OF METHOD

In what is known as the 'thematic apperception test' a series of somewhat vague and problematic situations are presented to the subject, and he is asked to make up a brief story explaining the situation, and indicating the feelings and thoughts of the principal characters. In the more usual form the material consists of standard pictures; in other forms the situations are described verbally. The psychologist then attempts to infer the personal characteristics of the individual tested from the nature of the 'themes' appearing in his stories.

The main object of the investigation described in the following pages was (i) to examine the possibility of scoring the pictorial form of the test in accordance with a more adequate theory of personality, and in particular (ii) to estimate the validity of assessments so obtained for certain basic traits.

In order to decide which traits are likely to lend themselves to assessment by such a test, it is essential to begin by asking what kind of mental activities such tests tend, in general, to provoke. This preliminary study has been strangely neglected by nearly all who have devised or used such procedures.[1]

The use of pictorial material for tests of character and personality has a far longer history than is commonly supposed. As long ago as 1886, Dr. Sophie Bryant,[2] the first headmistress of a famous London High School and one of Francis Galton's most active disciples, carried out what must have been the earliest set of experiments in testing character. As stimuli she used both pictures and words, and interpreted the answers chiefly in terms of the crude associationist theory which was then current (1). Binet, in his studies of individual differences, got children to describe both pictures and objects, and distinguished four main types—the observer, the describer, the erudite, and the emotional ((3) and other papers); later he included pictures in his scale of intelligence tests (4). Most of the older manuals of experimental psychology include experiments with ambiguous pictures, diagrams, or ink blots, designed to illustrate the principles of apperception;[3] and nearly all emphasize the importance of personal and emotional factors,[4] as well as tendencies not unlike those formerly described as 'primary and secondary laws of association.'[5]

[1] For much of what follows I am indebted to roneo'd Laboratory Notes by Professor Burt (15).

[2] Curiously enough, Vernon, in his survey of British work on personality tests, refers to none of the various papers published by Sophie Bryant. There are several references to her work in Whipple's account of 'tests of description and report' ((22), pp. 290, 292, 313, 319).

[3] E.g., Witmer, L., *Analytical Psychology: A Practical Manual* (1902), experiments 1–8; Seashore, C. E., *Elementary Experiments in Psychology* (1908), ch. xii.

[4] "Apperception," says Stout (and this is the special interpretation developed by British in contrast to continental writers), "is essentially a *conative* process; . . . it is the process by which a conation fulfils itself" (*Analytic Psychology*, II, 1896, pp. 114, 121): it is instructive to note how Stout's discussion of the 'conflict of apperceptive systems' (pp. 145–62) anticipates the Freudian doctrine of conflict, repression, and the influence of unconscious tendencies, and indicates their importance for the organization of personality.

[5] Cf. Stout, *op. cit.*, pp. 132f., on the 'conditions determining the strength of apperceptive systems.'

It was a growing realization of the defects of the older association theories that led British writers to develop the idea of a more general integrative principle : in doing so, their criticisms of associationism and their emphasis on organization rather than mere linkage largely anticipated the views of the Gestalt psychologists, particularly in the field of personality. The introduction of the term and idea of ' apperception ' into modern psychology is due chiefly to the Herbartian school, particularly in this country to Stout (2) and Adams (5). It may perhaps best be defined as " the process whereby the content of a particular sensory presentation, at first relatively indeterminate and meaningless, seems to become clearer and more determinate, and so acquires for the perceiver some kind of apparent *meaning*, in virtue of a fusion or synthesis with contents already systematically organized in the perceiver's mind." All Herbart's followers insist that a man's mode of apperceiving an object depends largely on his own individual ' apperception-systems,' and hence may incidentally provide a clue to his intelligence, interests, and character. Adams relates the legend of the Indian princess, who embroidered on her banner representations of herself, a coin of the realm, and her father's imperial crown, all superposed to make a single picture-puzzle ; each of her suitors was required to say what the device portrayed : the lover who saw, not the money or the crown, but the portrait won the contest. Steinthal, one of Herbart's best-known exponents, quotes a similar story : " we place a man," he concludes, " largely by the point of view from which he regards things, in short, by his mode of *apperceiving*."[1]

Binet, Stern, and most of the earlier investigators who used pictorial tests were interested chiefly in the examinee's ability to observe what is before him and to organize what he observes into an intelligent report. The first to exploit the principle of apperception for testing the dynamic aspects of personality appears to have been Burt. In describing his tests of ' emotional activities ' he writes : " Ambiguous pictures, that is, pictures which can be *apperceived* or attended to in various incompatible ways, are exhibited to the different individuals ; . . . and the particular mode of apperception shows consistent and significant differences." Some of the most conspicuous peculiarities were those revealing the distinctive interests of the two sexes (7). He observes that hitherto most writers had accepted Whipple's definition of a ' mental test ' as denoting ' the experimental determination of some phase of mental *capacity* ' ((6), p. 1), thus restricting it to cognitive processes ; the attempt to measure the affective and conative processes, or, as he sometimes terms them, the ' dynamic or orectic aspects ' of personality, has been " almost entirely neglected." He reports the results of various experimental procedures, and concludes that " the best tests of emotional disposition are to be obtained by the method of *impression* rather than of *expression*," i.e., by their influence on such processes as attention, association, preference, and apperception ((7), p. 381 ; cf. (9)).

Since the methods of scoring and interpretation adopted in the present investigation with adults were based on those worked out by Burt in his work with children, it may be helpful to describe quite briefly his general procedure and how the results were classified and marked.

In his re-standardization of the Binet intelligence-tests he notes how Binet's picture of a mysterious figure standing on a couch and looking out of a window elicited striking individual differences, quite apart from variations in observation or intelligence : " some children describe the man more or less objectively as he appears in the picture ; others subjectively identify themselves with the figure, and describe what he might be supposed to see—basing their descriptions largely on their own wishes, experiences, or fears." He accordingly retained it in his series of ' problem pictures.' The others used in these earlier researches were : Watts' *Hope*, Bramley's *A Hopeless Dawn*, Boecklin's *Toteninsel*, Millais's *Boyhood of Raleigh*, Yeames' *When Did You Last See Your Father ?*, Wallis's *Death of Chatterton*, Newcome's *What Shall He Be When Grown Up ?*, Dali's *Composition*, an illustration from the *Strand Magazine* entitled *She Was Looking at Me through the Half Open Door*, and two drawings from *Punch*. These twelve pictures were used as a standard series because they were readily accessible to others who wished to use the test. In later experiments he found that better results could be obtained by using specially designed pictures suggested

[1] Steinthal, H. (1881), *Einleitung in die Psychologie und Sprachwissenschaft*, pp. 167f. The whole passage, quoted by James, *Principles*, II, p. 108, and James' own discussion are well worth reading. James contends that apperception is merely a pretentious and somewhat ambiguous name for " the sum total of the effects of association." He would prefer G. H. Lewes' word ' assimilation ' (*Problems of Life and Mind*, 1879, I, pp. 118f., a discussion which emphasizes, with instructive illustrations, the individual nature of ' assimilations '). It was apparently James' objections to the term that caused its temporary disappearance from most American text-books.

by previous work and drawn by Miss Charleston ((16), p. 113). He describes the whole procedure as " resembling what older writers would have regarded as a ' free association ' test, with concrete scenes for stimuli instead of words " : " the visible form," he explains, " appeals more directly to primitive emotions and less to the higher critical feelings . . . whereas words lead more rapidly to emotional repression " ((10), pp. 326, 330).

The method of administration resembled that prescribed by Whipple for tests of description and report ((6), pp. 287, 301). In the group-form of the tests each child had a copy of the picture in front of him, and was asked first to write a free composition (' story ') and then to answer an ' interrogatory.' Titles were erased, and had to be supplied by the child. With younger children a roneo'd outline of a story was presented to be filled up like a completion test. In the individual form of the tests the story and the replies were given orally. The schedule for marking the results, however, instead of referring solely to ' accuracy and range of report ' (as in Whipple's version), included a set of headings for classifying and assessing replies, based on much the same scheme of personality as was used for other tests or assessments, noting (as had been done for the verbal association tests) the common types of clue to be found in the content of the responses and the apparent significance of each (21). The selection of clues was determined by a systematic item-analysis.

The quantitative results were only moderately successful : (average figures are, for reliability, 0·57, for validity, 0·16 to 0·38 : (16), p. 116). But the method proved to be, like the association test, a useful starting point which could be followed up in private interviews on quasi-psychoanalytic lines. It was therefore, as he explains, regularly used by himself and his co-workers in studies of delinquent and neurotic children.[1]

With the revival of associationism, owing to the influence of the behaviourist school,[2] interest in such concepts and procedures seems to have suffered an eclipse. There is no mention of apperception tests among the ' projection tests ' described in the first edition of Cattell's *Guide to Mental Testing*, nor yet in Earl's excellent review of ' methods of assessing temperament and personality.'[3] However, since the war the work of Murray and his colleagues at Harvard has led to widespread revival of interest in this type of test. To begin with, they apparently envisaged the method as a means merely for ' investigating fantasies ' (11), but later realized it could be used for more general purposes (12).

The type of material and method of interpretation developed by these writers is excellently described by Vernon (22), who also gives a critical and impartial summary of the chief results. " Binet, Burt, and others," he writes, " have used picture interpretation as a means of studying intellectual and emotional qualities, but the test in its present form was first described by Morgan and Murray : they collected a standard series of 30 photographs ; the subject is shown each picture, and asked to make up a story describing the situation, the events leading up to it, and the outcome, together with the thoughts and feelings of the characters. . . . Unfortunately the interpretation varies widely with the theoretical background of the psychologist. Those who follow Murray have adopted his elaborate system of needs and presses (external forces that the individual regards as beneficial or harmful) : Rapaport and others present simpler schemes. They are based on the style or structure —compliance with instructions, language characteristics, logical coherence, consistency of stories with one another, realisticness, etc., and on the content of recurrent themes. . . . Burt and Sen (in an unpublished memorandum) advocate a more straightforward approach, owing little or nothing to abnormal psychology.

" Many writers have testified to the value of T.A.T. in mental hospitals, clinics, etc., especially when used with other tests such as Rorschach. But scientific evidence is conflicting. . . . Harrison was able to assess I.Q.s with a validity of ·78. Sen finds the reliability of assessments by different scorers for Burt's qualities to average only ·4 ; nevertheless such qualities as Observation, Verbal, Level of Organization, and Maturity gave very promising correlations with follow-up results among high-grade civil servants " ((22), pp. 181–4).

[1] Cf. (9), (10), and *The Young Delinquent* (1925), pp. 404f.

[2] Even to-day, if a student, curious to know what is meant by an ' apperception test,' turns to the index of his text-book of psychology, he will find the word neither in Woodworth nor in Boring, Langfeld, and Weld. The temporary disappearance of the term in British literature is due largely to Spearman's influence. Referring to " apperception and assimilation," he maintains that his doctrine of the eduction of relations and correlates (really a revised form of associationism) " shows that neither term can supply anything that psychology needs " (*The Nature of Intelligence and the Principles of Cognition*, 1927, p. 26).

[3] Earl awards high praise to the inkblot test (as popularized by Rorschach and his followers)—a test which older writers showed was so largely dependent on apperception as above defined (14). Yet the intuitive methods of scoring the replies, as he describes, could only be accepted as plausible by one who was entirely unfamiliar with the theoretical and experimental work on apperception.

Several correspondents have since asked if fuller details could be made available in regard to both the method of approach here referred to and the results obtained. Accordingly, I have endeavoured to summarize as succinctly as possible the general principles adopted in using and marking the test, and the main conclusions so far emerging in what (it should be emphasized) are merely preliminary trials.[1]

II. SUBJECTS

The present investigation was based on the performances of 100 candidates for the Administrative Class of the Civil Service.

During the period of post-war reconstruction a new procedure was introduced by the First Commissioner (Sir Percival Waterfield), who also asked Sir Cyril Burt to act as Psychological Consultant to the Service and as a member of the Technical Advisory Committee. A full account of the general plan as finally adopted and of the tests and other methods employed has already been published by Mr. N. A. B. Wilson (19), who was at that time Principal Psychologist. Briefly it included a written qualifying examination, followed by a two- or three-day examination during a kind of ' house party ' by a Civil Service Selection Board (which included psychologists and was known as CISSB), and ending with an interview by a Final Selection Board (FSB). A mark or grading was given at each stage, but the ultimate decision rested with the Final Selection Board. This Board judged each applicant on his entire record, including the marks received in the earlier stages, and his prowess at the final interview (20). Detailed results indicating the apparent value of this procedure and of its component parts (excepting the projection tests) have been given by Professor Vernon (22).

The main criteria used in the present research were derived from reports sent in on successful candidates a year or more after their appointment. The reports were submitted on special ' follow-up forms ' containing spaces for (*a*) five-point ratings on the thirteen qualities believed to be most relevant to their duties, (*b*) an overall grading for general efficiency up to date, (*c*) an estimate of the highest rank in which the officer might be successful (termed the ' final rank '), and (*d*) a free pen portrait.

III. TEST MATERIAL AND MODE OF SCORING

Test Material. Neither the material used by Murray for American adults nor that used by Burt for school children would have been altogether suitable for candidates of this type. The pictures ultimately employed were the following : 1. A young woman surprised by an unexpected visitor ; 2. A boy holding a violin, with a girl standing by him ; 3. A man and a woman in conversation at a table ; 4. Two puppets ; 5. A woman sitting at a window, open book in hand, looking up at a man standing by her. Two-and-a-half minutes were allowed for the candidates to write a story about each.

At the time of my enquiry this set of pictures had not been in use long enough for a detailed item-analysis to be made of the various clues. The scheme of marking had consequently to be taken over very largely from work with somewhat different sets of pictures and a somewhat different type of examinee. Hence the whole enquiry can only be regarded as a tentative study of the applicability of the method proposed in a rather novel field.

[1] Mrs. M. M. McArthur (Principal Psychologist at CSSB) has very kindly read my draft ; and I am greatly indebted to her both for her co-operation in the experimental work and for her criticisms of my original report. I have also to express my thanks to the Civil Service Commission for permission to publish results obtained while I was working for the Selection Board. They and their officers, of course, have no responsibility for any of the views here put forward or for the conclusions reached.

Mode of Scoring. As Rapaport observes ((17), p. 423), the usual techniques for interpreting the thematic apperception test " are not hard-and-fast rules like those of scoring other tests ; they are rather viewpoints for looking on the stories, which have to become ingrained in the examiner." Murray and his collaborators drew up an elaborate scheme of ' needs ' and ' presses ' (which, as Cattell observes, is, in effect, an amalgam of McDougall's primary emotions with Freud's theory of fantasy formation) ; and " suggested and used depth interpretations which they compared with psychoanalytic interpretations of dreams " ((18), p. 214).

The immense number of headings involved in Murray's scheme of itself renders his method quite impracticable in any ordinary selection procedure ; and the peculiar terminology employed—' conservance,' ' infavoidance,' ' similance,' ' succorance,' and the like—is exceedingly difficult to interpret with any exactitude. Those who have endeavoured to improve upon it still choose their qualities or traits in the light of intuitive speculation or *ad hoc* improvisation, and so produce widely varying schemes for which there is really very little empirical support. Full details of the methods and of the modifications suggested by their followers are given by Rapaport and Bell, and summarized by Vernon. Vernon repeatedly stresses the extreme subjectivity of all such procedures ; and rightly adds that " the lack of any generally accepted framework for the description of personality creates further difficulties " ((22), p. 205). In addition it is hard to discover any precise information regarding either the reliability or the validity of the methods used by earlier investigators.

During their work with school children, extending over more than forty years, Burt and his co-workers have evolved a structural scheme for the study of personality ; and in their researches on pictorial tests they have tentatively worked out a systematic procedure for analysing both the formal and material characteristics of the narrative compositions of children (9). It has been plainly influenced by the general practice of literary and biographical critics ; but differs in retaining the same scheme for all the writers to be studied, in attempting to give a quantitative estimate for each characteristic, and above all in having been subjected to some degree of empirical verification. It employs a stock schedule for the classification of the commoner *motifs* (recurrent topics, subjects, or themes) according to the dominant emotions and salient interests (sentiments and complexes) which each appears to indicate, with special reference to their dynamic interactions.

The choice of traits has been based on the results of factorial studies ((8), (13), (16)), and checked by systematic comparisons with the after-histories of the cases tested ; the inferences drawn can thus be fitted into a recognized psychographic scheme available for other tests. It is often objected that assessments for qualities derived by factorizing tests or traits fail to give a true synthetic portrait of the individual as a whole. To overcome this difficulty, Burt also grades the traits in order of their prominence in one and the same person. This enables a factorial ' psychogram ' to be constructed, and yields a sounder estimate for what are sometimes called ' temperamental types ' and should perhaps rather be termed ' temperamental tendencies.' [1]

With a given set of pictures the commoner features of the stories which they tend to elicit are arranged according to the various cognitive or orectic characteristics they are presumed to indicate. In marking the scripts, a prepared schedule is generally used ; pluses are entered in the appropriate spaces for the several signs or symptoms noted in each candidate's script (pluses or minuses in the case of bipolar traits). An allowance is given for over-all impressions. These scores are then added up, and treated as measurements for the several traits.

No doubt with pupils from the elementary schools the structure of the individual's personality is much simpler than that of the highly intelligent adult. Nevertheless, it seemed profitable to enquire whether this scheme might not provide a better starting-point than the unverified speculations which seemed to form the basis of the American techniques. In the investigations described below, an attempt was made to score the results of the test in accordance with these principles. The scheme covers both cognitive abilities and orectic (affective and conative) characteristics : but for practical reasons it could not be adopted in its entirety. In the present enquiry the figures quoted will be confined to those particular characteristics which

[1] Cf. (13), and *Factors of the Mind*, p. 427, Fig. 1.

seemed to be most closely related to the duties that would be required of the successful candidates and which previous attempts had shown could be rated with greatest reliability.

IV. RELIABILITY AND VALIDITY OF ASSESSMENTS

A. *Reliability.* The stories written by a hundred candidates were scored by Mrs. McArthur and myself. Further, the scripts of twenty candidates at an earlier examination, which I had already scored in the same way, were re-marked after an interval of nine months, when, owing to intervening work, I found I had completely forgotten my original impressions. The results obtained (*a*) for the self-consistency of the same interpreter and (*b*) for the reliability of the two independent assessments are shown below in Table I. With a sample of 100, all product-moment coefficients over 0·197 and 0·256 are significant at the 5 per cent. and the 1 per cent. levels respectively ; with a sample of 20, coefficients above 0·444 and 0·561 are significant at those levels.

TABLE I. COEFFICIENTS OF SELF-CONSISTENCY AND RELIABILITY

Traits Assessed	Self-consistency of One Judge	Reliability of Two Judges
A. *Cognitive*		
1. General Organization of Story	(assessed only once)	
2. Observation	·492	·287
3. Verbal Ability	·758	·467
4. Imagination	·605	·588
B. *Orectic*		
5. General Emotionality	·634	·395
6. Extraversion–Introversion	—	·222
7. Cheerful–Depressive	·715	·628
8. Self-assertive–Self-submissive	·426	·466
9. Anxiety	—	·451
10. Decisiveness	·560	·418
11. Ambition	·169	·295
12. Social Consciousness	·804	·452
13. Maturity–Immaturity	·718	·485
14. Integration	—	·371
Average	·588	·425

If one may judge from the more careful studies reported in the literature, the reliability of so-called ' projective tests of personality ' seldom reaches a figure of 0·400 ; and this was approximately the average value in our previous attempts. The table above shows that all except one of the self-consistency coefficients, and nine of the reliability coefficients, rise above that level. Somewhat unexpectedly the quality which is most consistently assessed is Cheerfulness. The three poorest reliabilities which fall well below the figure of 0·400 are those for Observation, Ambition, and Extraversion–Introversion : the last indeed is non-significant. Further study indicated that, in these cases, the low coefficients were due to the fact that the two investigators seemed often to be looking for slightly different things : e.g., in assessing Observation, one of the judges looked for details regarding physical appearance, dress, gesture of the characters in the story (i.e., indications of observation in general life), whereas the other interpreted the term more narrowly as meaning accurate observation of what actually appeared in the picture. In this and other instances much of the discrepancy arose from an inadequate definition of the quality to be assessed and from incomplete

specifications of the relevant clues. The term Ambition, for example (described as ' an ardent desire for personal distinction or glory '), is a somewhat ambiguous concept ; and the phrase ' Self-assertion *versus* Self-submission ' (described as McDougall defines those tendencies) appears to treat aggressive tendencies and self-assertive tendencies as almost synonymous : although correlated, it seems clear that the two should be assessed separately. Moreover, some of the pictures were not altogether suited to elicit trustworthy clues to the qualities required. But these are defects in the procedure which could doubtless be corrected after further research.

B. *Validity.* In judging validity the criteria used were (*a*) the marks allotted by the Final Selection Board after the three-day examination and (*b*) the various marks and gradings provided by the follow-up forms. Data sufficiently complete for comparison were available for only 42 of the successful candidates. The scores for all the traits were first correlated with suitability as estimated (i) by the Final Selection Board, (ii) by the Over-all Grading, and (iii) by the estimate of ' Final Rank.' Table II shows the results.

With samples of this size, correlations above 0·304 and 0·393 are significant at the 5 per cent. and 1 per cent. levels respectively. The correlations are no doubt greatly reduced by the fact that those candidates who obtained low marks for desirable qualities failed to secure appointment ; they had consequently to be omitted from the calculations for validity. Any attempt to correct for this type of selection by statistical calculation seemed highly precarious. Hence the figures are printed as they stand, and must be compared with uncorrected coefficients given in other reports, not with coefficients already corrected for selection.

TABLE II. CORRELATIONS WITH SUITABILITY

Traits Assessed	(i) Final Selection Board	(ii) Over-all Grading	(iii) Final Rank	Average
A. *Cognitive*				
1. General Organization of Story ..	·427	·318	·354	·367
2. Observation	·268	·331	·401	·333
3. Verbal Ability	·344	·237	·359	·313
4. Imagination	·328	·168	·254	·250
B. *Orectic*				
5. General Emotionality	·127	·100	·203	·143
6. Extraversion–Introversion	·219	·118	·080	·108
7. Cheerful–Depressive	·117	·147	·183	·149
8. Self-assertive–Self-submissive ..	—·070	—·228	—·151	—·150
9. Anxiety	—·149	—·130	—·104	—·128
10. Decisiveness	·198	·074	·137	·136
11. Ambition	·029	·044	·027	·033
12. Social Consciousness	·037	·094	·207	·113
13. Maturity–Immaturity	·309	·183	·335	·276
14. Integration	·233	·299	·325	·286
Average	·204	·170	·231	·201

NOTE. *In Tables II and III the negative signs have been reversed in calculating the averages at the foot of the columns.*

Owing to the small size of the sample, only six traits furnish one or more correlations that are statistically significant, and of these four are cognitive. Nevertheless, the figures are on the whole higher than the validities obtained for most of the selection traits used at CISSB (4), where the highest (uncorrected) correlation with the criterion was 0·186 (for the ' Short Talk '). In general the estimate

of Final Rank correlates more closely than the other two. The highest average correlation is obtained by the marks for ' General Organization of Story.' This is the only characteristic that provides a fully significant coefficient with all three gradings : it may perhaps be taken as a measure of the kind of verbal and logical intelligence that is required in such candidates. The Final Selection Board mark correlates more highly than the other estimates with the scores for Organization of Story, Imagination, Lack of Anxiety, Extraversion, and Decisiveness. The estimate of Final Rank, which gives a long-term view of a candidate's suitability, correlates most highly with Observation and Verbal ability among the cognitive qualities, and with Maturity, Integration of Personality, Social Consciousness, Emotionality, and Cheerfulness among the orectic. The former group of qualities are in the main those which we might perhaps expect to tell most at a personal interview ; the latter, those which on the whole might be expected to make for general efficiency of work in a government department.[1]

Since different Government departments require slightly different qualities, an attempt was made to determine how far these more specific qualities, as assessed by the supervisors, could be predicted by the marks for the several traits that approximate most closely to these qualities. Table III shows the correlations between the most relevant assessments derived from the test and the nearest assessments appearing on the report forms. All the correlations point in the right direction. But the only correlations that are significant at the 5 per cent. level are those for Observation and Imagination.

TABLE III. CORRELATIONS FOR SPECIFIC TRAITS

	(*a*) Traits Assessed by Test	(*b*) Traits Assessed on Follow-up Form	Correlation
2.	Observation	Attention to detail	·304
3.	Verbal Ability	Paperwork (minutes and correspondence)	·256
4.	Imagination	Fertility of ideas	·384
6.	Extraversion–Introversion	New contacts	·281
6.	Extraversion–Introversion	Acceptability as colleague	·116
7.	Cheerful–Depressive	Judgement	·196
8.	Self-assertive–Self-submissive	Acceptability as colleague	—·260
8.	Self-assertive–Self-submissive	Management of subordinate staff	—·204
10.	Decisiveness	Judgement	·264
		Average	·252

So far as the foregoing results take us, it may, I think, be reasonably claimed that the analytic method here adopted for evaluating the results obtained with a pictorial apperception test is unquestionably the most satisfactory for securing objective evidence regarding the reliability and validity of such procedures. It would be instructive to compare the over-all picture that might be derived for each candidate from the detailed assessments with the results of the synthetic judgements submitted by those who prefer to rely on intuitive impressions. What, however, seems most urgently needed is a further improvement of the details of the technique. Probably it would be wiser to begin first by selecting a more informative set of pictures in the light of the experience already gained. After that a systematic item-analysis should be undertaken to determine first those traits that can be most reliably assessed by this particular method and then what detailed clues obtained with the pictures selected yield the most valid judgements.

[1] A factor analysis of the follow-up form (20) showed that the Over-all Grading falls on the same side of the bipolar analysis as the cognitive qualities (such as knowledge of work, fertility of ideas, quick apprehension, and attention to detail), whereas Final Rank is grouped with the orectic or ' personal ' qualities (long-term contacts, new contacts, and effectiveness as member of a committee).

V. SUMMARY AND CONCLUSIONS

1. The responses to a thematic apperception test obtained from a hundred Civil Service candidates in the course of the regular entrance examination were scored by two investigators according to Burt's method of rating for a systematic scheme of traits based primarily on the results of earlier factorial analyses. The correlations between separate assessments made by the same judge at an interval of nine months showed a moderately high consistency for nearly all the traits (average 0·59). The correlations between the independent ratings made by two judges showed that, when compared with similar coefficients for other projection tests of personality, or other methods of marking, the assessments obtained by this method possessed a degree of reliability that is encouraging, though still not high enough to have much practical value as they stand. Three traits at least seem to be ill-defined and possibly may not lend themselves to assessment by this type of test. The traits yielding the largest reliability coefficients were Verbal Ability, Imagination, Cheerful or Depressive tendencies, Self-assertive or Self-submissive tendencies, Anxiety, Decisiveness, Social Consciousness, and Maturity versus Immaturity.

2. The validity coefficients are higher than those obtained with other tests of the same characteristics and other methods of marking, and are superior to those hitherto published for similar assessments of Civil Service candidates : the correlations between the assessments based on the test and the various criteria available suggested that, of the traits assessed, those most closely related to suitability for administrative work appeared to be Ability for Coherent Organization of the Story, Observation, Verbal Ability, Imagination, Maturity, Cheerfulness, and Integration of Personality. Correlations with reports on the successful candidates' efficiency sent in a year or so after their appointment indicated that several of the more specific traits could be assessed with a promising degree of accuracy.

3. The enquiry as a whole indicates that both the schedule of traits and the method of scoring form a distinct improvement on those adopted by previous British workers (for the most part borrowed, with some simplification, from American contributions). At the same time the detailed results reveal that certain features in the procedure leave considerable room for improvement ; and a study of the chief sources of discrepancy has suggested what will probably be the most fruitful directions for future research. It seems clear that the efficacy of the thematic apperception test can be greatly improved if the traits to be assessed are based on an acceptable theory of personality-structure, and if the various clues used in the marking-scheme are selected on the basis of a systematic item-analysis and are explicitly specified and defined.

REFERENCES

1. Bryant, S. (1886). ' Experiments in testing the character of school children.' *J. Anthrop. Inst.*, XV, 338–49.
2. Stout, G. F. (1888). ' The Herbartian Psychology.' *Mind*, XIII, 321–38, 473–98 ; XIV, 1–26, 353–68.
3. Binet, A., and Henri, V. (1896). ' Psychologie individuelle.' *Ann. Psychol.*, III, 296–332.
4. Binet, A., and Simon, T. (1908). ' Le developpement de l'intelligence chez les enfants.' *Ann. Psychol.*, XIV, 1–94.
5. Adams, J. (1910). *The Herbartian Psychology applied to Education.* London : Heath.
6. Whipple, G. M. (1910). *Manual of Mental and Physical Tests.* Baltimore : Warwick and York.
7. Burt, C., and Moore, R. C. (1912). ' The Mental Differences between the Sexes.' *J. Exp. Ped.*, I, 273–84, 355–88.

8. Burt, C. (1915). 'General and Specific Factors underlying the Primary Emotions.' *Brit. Ass. Ann. Rep.*, LXXXIV, 694–6.
9. Burt, C. (1934). 'The Interpretation of Pictures.' Unpublished Laboratory Notes. University College, London.
10. Burt, C. (1935). *The Subnormal Mind.* London: Oxford University Press.
11. Morgan, C. D., and Murray, H. A. (1935). 'A Method for Investigating Fantasies: the Thematic Apperception Test.' *Arch. Neurol. & Psychiat.*, XXXIV, 289–306.
12. Murray, H. A., *et al.* (1938). *Explorations in Personality.* New York: Oxford University Press.
13. Burt, C. (1939). 'The factorial analysis of emotional traits.' *Character and Personality*, VII, 1–238, 285–99.
14. Earl, C. J. C. (1939). 'Methods of assessing temperament and personality,' *ap.* Bartlett, F. C., *et al. The Study of Society* (ch. X). London: Kegan Paul.
15. Burt, C. (1945). 'The Thematic Apperception Test.' Unpublished Laboratory Notes. University College, London.
16. Burt, C. (1945). 'The Assessment of Personality.' *Brit. J. Educ. Psychol.*, XV, 105–21.
17. Rapaport, D. (1946). *Diagnostic Psychological Testing.* Chicago: World Book Company.
18. Bell, J. E. (1948). *Projective Techniques.* New York: Longmans, Green & Co.
19. Wilson, N. A. B. (1948). 'The Work of the Civil Service Selection Board.' *Occup. Psychol.*, XXII, 204–12.
20. Vernon, P. E. (1950). 'The Validation of Civil Service Selection Board Procedure.' *Occup. Psychol.*, XXIV, 75–95.
21. Sen, A. (1950). 'A statistical study of the Rorschach test.' *Brit. J. Psych., Stat. Sect.*, III, 21–39.
22. Vernon, P. E. (1953). *Personality Tests and Assessments.* London: University of London Press.

Vol. VI
Part II

The British Journal of Statistical Psychology

November,
1953

THE FACTOR ANALYSIS OF MATRICES OF NEGATIVE CORRELATIONS

By PATRICK SLATER
Bedford College, University of London

I. *Introduction.* II. *Statistical Conditions for Negative Matrices.* III. *Experimental Conditions.* IV. *Analysis of New Example.* V. *Analysis of the Previous Example.* VI. *Methodological Discussion.* VII. *Summary.*

I. INTRODUCTION

In the discussion of a matrix of predominantly negative correlations in a previous paper (1), a first attempt was made to define the general conditions under which such matrices might occur. Taking into consideration another instance of a slightly different type (cf. Sect. III), a more adequate definition can now be reached.

In the earlier paper only a few incidental comments were made on the possibility of applying factor analysis to such matrices. It was mentioned that an imaginary general factor (i.e., one with coefficients of $i = \sqrt{-1}$ for its loadings) might be obtained, if a centroid analysis were applied mechanically, and that a plausible psychological explanation might even be given for such a factor, since a particular range of mental processes (involving an element of choice) would be found in most psychological experiments where negative matrices were likely to occur. But it was shown that the need for extracting an imaginary general factor could be eliminated by introducing a transformation.

Burt, in a *Reply* in the same issue (2), expressed the view that a factor analysis would be possible without the extraction of any imaginary factor, and he doubted whether any preliminary transformation was required. He gave a factor analysis of my figures, but did not include a complete answer to the question discussed in detail here, how factor analysis should be applied to such matrices. However, the main conclusion now reached is in close agreement with Burt's original opinion. It is in fact shown that an efficient analysis is possible without transforming the matrix or extracting any imaginary factor, a transformation being only required in the test of significance.

Both the original example and the one adduced here are instances of matrices of variances and covariances of variables in a common metric, which do not need to be standardized. This permits the use of a variant of Bartlett's test ((3), (4), and (5)) which has theoretical advantages but is appropriate only when factor analysis is applied without prior standardization. In experiments involving ratings, rankings, and comparisons, where negative matrices are particularly likely to occur, a common metric can frequently be used for recording the simultaneous variations ; and an illustration of the treatment of unstandardized matrices may be found helpful.

II. STATISTICAL CONDITIONS FOR NEGATIVE MATRICES

Suppose a matrix of correlations has been obtained by measuring m variables, A, B, etc., in n cases, the observed measurements in a particular case being $a_0, b_0 \ldots m_0$. The sum of these measurements, s_0 say, will usually also vary from case to case, and may be treated as a further variable, S. If V_a, V_b, etc., denote the total variances of the variables and W_{ab}, etc., their covariances, we may write $\Sigma V = V_a + V_b + \ldots V_m$, and $\Sigma W = W_{ab} + \ldots W_{lm}$ (summing the covariances without duplication) and shall then obtain, for the total variance of S, $V_s = \Sigma V + 2\Sigma W$, i.e., the sum of all the terms in the square matrix of total variances and covariances. When ΣW is negative, i.e., when $\Sigma V > V_s$, the matrix of correlations will be preponderantly, though not necessarily completely, negative.

This may happen if the conditions of the experiment restrict the variation of S directly, while only indirectly restricting A, B, etc. If S is completely restricted, $V_s = 0$, and consequently $\Sigma W = -\frac{1}{2}\Sigma V$. This is evidently the negative limit of ΣW when A, B, etc., are real scalar measurements. For if it exceeds this, V_s must be negative ; but as V_s is a sum of squares, it cannot be negative unless some of its component terms, such as s_0, are imaginary or complex ; and as s_0 too is a sum, this implies that some of its components, a_0, etc., must be imaginary or complex also. In short, some at least of the dimensions in which variation has been recorded must have been imaginary or complex *ab initio*. Matrices of this kind are not discussed below, for lack of any experimental example. But there would not appear to be any objections in principle to extracting imaginary factors from them.

Where $V_s = 0$, the conditions of the experiment must have excluded variation in S ; and thus the matrices obtained resemble matrices from which a variable has been partialled out statistically. Methods for their analysis can be adapted from methods appropriate for bipolar matrices of residual correlations. Difficulties in their treatment arise because their determinants are zero, and because in the absence of association the correlation between pairs of variables does not tend to 0 but to $-1/(m-1)$. Tests of significance are complicated accordingly. Degrees of freedom and numbers of latent roots also require special attention.

When V_s is positive but less than ΣV, the matrix may be unusual in superficial appearance only ; but the difficulties which are encountered when $V_s = 0$ must become gradually more obtrusive as it approaches this limit. The value of the determinant must recede towards vanishing point, and the null value of the correlation coefficient shift from 0 towards $-1/(m-1)$.

The method of principal components can be applied to either type of matrix without postulating any imaginary factor. When all the observed variance is in real scalar dimensions it would be difficult to find valid grounds for inferring the occurrence of variation in any imaginary dimension. And the extraction of an imaginary factor would leave the residual variances of the variables greater than their original totals. Generally, therefore, to postulate such an imaginary factor would appear unnecessary, improper, and absurd.

III. EXPERIMENTAL CONDITIONS

A simple instance of a negative matrix resulting from a partial restriction of V_s is provided by the following data from an investigation into the personal histories of two groups of men in training for an operational duty. One group consists of men already in the service posted on to the course from their previous units ; the other of men who have re-entered as volunteers after completing a previous period of service and returning to civil life for a time.

The number of years they had completed since reaching the age of 14 in (E) full time education (outside the service), (C) civilian employment, and (F) service (including technical training) were recorded, these being the only ways in which substantial parts of their lives had been spent. Since

the age for admission to the duty is restricted, those who have occupied more time in one way will have passed less in another ; and the three variables are negatively correlated. Table I gives the main statistics.

TABLE I. ANALYSIS OF DATA FROM MEN IN THE FORCES

Variable	A. *Means*		B. *Mean Square Variances and Covariances* ($d.f. = 69$)			C. *Correlations*		
	Transferred ($n = 54$)	Re-entrants ($n = 16$)	E.	C.	F.	E.	C.	F.
(E) Education	1·65	2·25	1·5912	−·9213	−1·3582	—	−·3340	−·4078
(C) Civil employment	1·39	4·19	− ·9213	4·7818	− ·6431	−·3340	—	−·1114
(F) Service	5·13	3·19	−1·3582	−·6431	6·9723	−·4078	−·1114	—

When the simultaneous variation of a number of variables is subject to the condition that some cannot increase unless others diminish, a choice between the variables must occur at some point. This may not always be the deliberate choice of a human being, but in psychological experiments it often is. Thus in this experiment an element of choice may be found in the way in which each man decided to employ his time : the statistics represent the accumulated effects of such acts of decision. It may also be found in the alternatives open to the selection board, who cannot simultaneously require extensive experience in many different fields if they wish to recruit men still in the prime of life.

In psychological experiments with methods of ranking or paired comparison the entire variation is usually attributable to acts of deliberate choice. In preferring one of the alternatives presented to him the subject declines another. Correlating preferences for different alternatives thus gives rise to predominantly negative matrices. Face to face with death there is not much deliberate choice open to an individual, but the situation is analogous in that by dying in one way he escapes dying in any other. A choice is made for him if not by him. Thus negative correlations are commonly found when fluctuations in mortality from different causes are correlated over a period during which the numbers and age composition of the population are tending to remain constant.

If such analogies are admissible, it would seem reasonable to enquire after an element of choice in the experimental conditions whenever a negative matrix is observed. Conversely, when an experiment involves some such element, some approximation to a negative form of matrix may be expected, and the method of analysis adopted should take this into account.

The mortality statistics illustrate another point. If the period under consideration is one when the numbers and age composition of the population are undergoing substantial changes, years when deaths from any particular cause are common are likely to be years when the total number of deaths is large ; and a general tendency towards positive correlation may appear among the causes. In the former case the total number of deaths from all causes (S in the notation of Sect. II) is not free to vary widely ; in the latter it is. The transition from predominantly positive correlation matrices, yielding positive general factors of the familiar kind, to predominantly negative ones, productive (when improperly analysed) of imaginary general factors, may thus be seen to be gradual and subject to control by any processes or conditions which expand or contract V_s relatively to ΣV.

IV. ANALYSIS OF NEW EXAMPLE

When a three-dimensional model of the total dispersion matrix in Table I B was constructed, no startling peculiarities were found. Its axes, computed by the method of principal components, were found to be as shown overleaf. Since the variables were expressed in a common unit (number of completed years) they were left unstandardized.

The results may be compared with those obtained by a discriminantal analysis. The function maximizing the discrimination between the two groups requires the relative weights : (E) −1·3465, (C) −1·3151, (F) + ·1270 ; 46·25 per cent. of its total variance is attributable to the differences between the groups. The determinants of the matrices of variances and covariances, total and within groups, have the corresponding ratio, 11839632 : 6365674.

Variable	*Factor loading* First	Second	Third
(E) Education	.. — ·5422	.. — ·6668 ..	·9234
(C) Civilian employment	.. — ·4606	.. 2·1213 ..	·2638
(F) Service	.. 2·6193	.. ·2360 ..	·2369
Latent root	.. 7·3667	.. 5·0004 ..	·9783

In internal factor analysis it is generally assumed that if the individuals measured differ in some respect, and the measurements recorded are sensitive to the difference, the major dimensions along which the individuals are found to vary may be used to infer and define the nature of the difference between them. This is the basic factorial hypothesis popularly translated into a belief that factor analysis can be used to prove or disprove the existence of psychological traits.

The major axis of the total dispersion matrix should accordingly approximate to the axis along which the dispersion between the groups is greatest. Actually the two axes diverge considerably. The major amount of variation between the individuals is in their length of service (cf. the first factor, above) which involves, as it increases, a diminishing length of time spent on education or in civilian employment. The variation between the groups, however, is mainly due to differences in the amount of time they have engaged in full time education or civilian employment, especially the former (the discriminant weights need to be related to the mean square variances, Table I B). Pooling square methods may be used to show that the two axes correlate by ·57, i.e., lie at an angle of approximately 55° to one another. The difference between the groups contributes only 14·6 per cent. of the total variance of the first component, and slightly more, 18·1 per cent., of the second.

V. ANALYSIS OF THE PREVIOUS EXAMPLE

The data given in the previous paper were obtained by asking five groups, saleswomen (S), light factory workers (F), teachers (T), nurses (N), and clerks (C), 18 women in each group, to score the five occupations for smartness of appearance off duty by the method of paired comparisons. Here s_0, the sum of the scores assigned by each subject, was constant and consequently $V_s = 0$.

The matrix of mean square variances and covariances is shown in Table II A below (obtained from the sum of the two matrices given in Tables III A and III B of the original paper). The principal components are given in Table II B.

TABLE II. ANALYSIS OF DATA FROM EMPLOYED WOMEN

Variable	A. *Variances and Covariances* (*d.f.* = 89) S.	F.	T.	N.	C.	B. *Principal Components* *Factor loading* First	Second	Third	Fourth
S ..	5·0164	·8904	−3·7686	−1·8382	— ·3000	−1·9044	·4136	— ·2329	1·0791
F ..	—	3·2265	−2·2000	−1·6427	— ·2742	−1·0779	— ·2852	1·2196	— ·7041
T ..	—	—	6·0270	·7034	— ·7618	2·1845	— ·7602	·3727	·7336
N ..	—	—	—	4·2517	−1·4742	1·0983	1·6652	— ·2513	— ·4576
C ..	—	—	—	—	2·8102	— ·3005	−1·0334	−1·1081	— ·6512
Latent root	—	—	—	—	—	10·8572	4·7613	2·9716	2·8319

Teaching is associated with nursing at one end of the major axis of variation, and sales with factory work at the other ; the diversity of opinion about the relative status of the professions is familiar and may be treated as a factor. The first two factors have loadings comparable with those

obtained by Burt (2) with a different procedure. The analysis concludes with the extraction of four factors, the fifth having vanished.

The two major components (λ_1 and λ_2) coincide closely with the two principal axes (μ_1 and μ_2) of variation between groups, previously published. Their inter-correlations are as follows :

	μ_1	μ_2
λ_1	—·9261	·2998
λ_2	·3733	·8055

The basic factorial hypothesis (that the difference admitted by the sampling will be located by the analysis) attains a much greater validity here.

VI. METHODOLOGICAL DISCUSSION

Factor analysis is usually applied in experiments in which variation has been recorded in a relatively large number of original dimensions, with the intention of showing that a smaller number of derived dimensions will account for most of it. The dimensions in which variation is greatest are found first (the method of principal components being the most efficient) ; and the remainder may be presumed to be dispersed by chance approximately equally among the other dimensions. Bartlett uses this hypothesis for his test of significance.

Without modifying the intention of the experiment, the introduction of the restriction described above modifies the outcome. As V_s diminishes towards 0, the dispersion contracts in one of its m dimensions, and finally collapses into a hyperplane of $m - 1$ dimensions. The latent root thus reduced recedes to last in order of extraction before ultimately disappearing. Thus in the first example the last latent root is relatively small ; and in the second it has gone.

It seems doubtful whether Bartlett's test is suitable when V_s is partially restricted. As the significance of each successive component is judged by comparison with the remainder, a disproportionate reduction in one of the roots must enhance the apparent significance of its predecessors. When it has vanished, this difficulty disappears.

The test assumes its simplest and most general form when applied to a matrix of unstandardized variables, such as Table II. After k roots have been extracted the reduced χ^2 needed for testing the significance of the remainder is

$$\chi^2 = -\left\{n - \frac{1}{6}(2p + 1 + 2/p) - \frac{2}{3}k\right\} \log_e V_{p-k},$$

with $\frac{1}{2}(p - k - 1)(p - k + 2)$ d.f. Here n is the d.f. of the variables, and p the total number of non-zero latent roots (i.e., $m - 1$ in the notation of Sect. II). V_{p-k} corresponds to R_{p-k} in (3), the latent roots of the unstandardized matrix being scaled down to an average of 1.

For the additive table of χ^2 in Table III the coefficient of $\log_e V_2$ has been used for all the values of V_{p-k}. The first latent root is the only one that appears to be beyond the limits of chance.

TABLE III. SIGNIFICANCE TESTS

Latent Root	*Analysis of* χ^2 d.f.	χ^2	*Values of* V_{p-k}
λ_1 2·0273	4	48·90	V_4 ·52884
λ_2 ·8890	3	7·59	V_3 ·91747
λ_3 ·5549	2	·05	V_2 ·99942
λ_4 ·5288	—	—	—
Total	9	56·54	—

If the test is to be applied before all the latent roots have been extracted, their mean, needed to scale them down, can be found from $\Sigma V/p$, and their product, needed for V_p, from $|T'MT|$. Here M is the matrix to be analysed, ΣV is the sum of the terms in its leading diagonal, p the number of latent roots, and T the transformation matrix quoted in the previous paper, with transpose T'.

Such a transformation may thus be used for the test, but is not needed for the analysis. When the components are extracted directly from M, their loadings are related directly to the original dimensions, and are thus easier to interpret. The use of T is therefore not recommended except for the test.

If a disproportionately small latent root is suspected as a result of a restriction on V_t, it could be eliminated by partialling out the variance of S; and the analysis would then follow the second model. To treat the matrix in Table I B in this way, chronological age would be held constant. But such treatment might not be suitable or even clearly intelligible in every instance.

The factorial hypothesis would not ordinarily be required when the origin of the heterogeneity in the sample is known independently, as in these examples. It is more likely to be helpful when it is known that the sample is heterogeneous in some respects but the individuals composing it cannot be classified independently. However, its application to such samples as these illustrates its validity, which is of course limited and varies from one experiment to another.

VII. SUMMARY

1. A matrix of negative correlations is defined here as one derived from exclusively or at least predominantly negative covariances. Such matrices can occur even when the variables concerned yield only real scalar measurements, provided the sum of the measurements in each case can vary only to a limited extent if at all. They are particularly likely to occur in experiments fixing a limited number of alternatives between which a choice has to be made or imposing an analogous set of conditions. The transition from a predominantly positive to a predominantly negative matrix occurs gradually as the limitations become more rigid.

2. To postulate an imaginary general factor in analysing them is unnecessary and confusing: their principal components can be found directly; no transformation is required. But the fact that as the transition from positive to negative occurs one of the latent roots is reduced and finally eliminated is important for factor analysis. If it is eliminated, a test of significance can be applied; and here a transformation may be useful. No direct solution is offered for the problem of testing significance when it is merely reduced, but it may be avoidable in some instances by partialling the root out completely before proceeding to analyse.

3. Circumstances in which factor analysis might be applied have been described; and certain problems involved in determining the psychological connotations of factors, and in assessing their validity, have been briefly discussed in connexion with certain concrete examples.

REFERENCES

1. Slater, Patrick (1951). 'The transformation of a matrix of negative correlations.' *Brit. J. Psych., Stat. Sect.*, IV, 9–18.
2. Burt, C. (1951). 'Reply to Mr. Slater's Criticisms.' *Brit. J. Psych., Stat. Sect.*, IV, 18–20.
3. Bartlett, M. S. (1950). 'Tests of significance in factor analysis.' *Brit. J. Psych., Stat. Sect.*, III, 77–85.
4. Bartlett, M. S. (1951). 'A further note on tests of significance in factor analysis.' *Brit. J. Psych., Stat. Sect.*, IV, 1–2.
5. Bartlett, M. S. (1951). 'The effect of standardization on a χ^2 approximation in factor analysis.' *Biometrika*, XXXVIII, 337–44.

Vol. VI Part II The British Journal of Statistical Psychology November, 1953

NOTES AND CORRESPONDENCE

ON THE OBLIGATION TO REFUTE RIVAL HYPOTHESES

By L. F. RICHARDSON

In the recent controversy between Dr. H. J. Eysenck and his reviewer (W. L. G.), much emphasis is laid by the latter on the need to disprove alternative hypotheses (this *Journal*, V, pp. 208f.; VI, pp. 44f.). I am impressed both by the importance of this requirement and by the difficulty of satisfying it. The practical difficulty is obvious: for how can anybody be certain that he has thought of all the possible alternatives? They can be so extremely diverse. Take, for example, a long-lasting theory, that of Sir Isaac Newton about the motion of the planets. At the time when it was first discussed, permit me to imagine that there may have been people whose alternative was: "God created the planets; the planets obey their Creator; this view of their movement is comprehensive and satisfying; Mr. Newton's notions about attraction and momentum are superfluous." Even to-day it would be difficult to refute that alternative: all we can say is that it yields no predictions. In the next century Clairaut's alternative theory was that the attraction did not vary precisely as the inverse square, but as some slightly different function of the distance (H. Lamb, *Dynamics*, Cambridge, 1920, p. 271). From the year 1916 the much-discussed alternative theory was Einstein's, and it was strangely different from any alternative that had been proposed in the previous two centuries. Sir Harold Jeffreys, after mentioning some small discrepancies between theory and observation that still remain unexplained, says: "There has not been a single date in the history of the law of gravitation when a modern significance test would not have rejected all laws and left us with no law. Nevertheless the law did lead to improvement for centuries . . ." (*Theory of Probability*, 2nd ed., p. 362).

As a consequence of W. L. G.'s insistence, I try now to refute the alternatives to my theory of arms-races (this *Journal*, current number, p. 89). Very briefly my own theory is: "The present arms-race is due to mutual stimulation." What are the alternative theories? The theory prevalent in Britain and U.S.A. is: "The present arms-race is entirely caused by the aggressive policy of Russia and her satellites." Another theory, prevalent in Communist countries, is that "the present arms-race is entirely caused by the aggressive designs of the capitalist countries." I do not indeed feel able to refute these theories one at a time: but do they not refute each other, at least as far as the assertion 'entirely' is concerned? What is left of each, when 'entirely' is omitted, is the mutual stimulation which I have formulated.

REPLY BY THE REVIEWER

I am glad that so experienced a scientist as Dr. Richardson has taken up the questions raised by the recent enthusiasm of certain psychologists for a purely 'hypothetico-deductive methodology,' and that he agrees, at any rate in principle, with what I said about the need to refute alternative hypotheses. I willingly accept his main criticism of my argument, namely, that it failed to recognize the practical difficulties involved in trying to satisfy the logical requirements. If I did not explicitly refer to this point, that was partly because my reply to Dr. Eysenck was already too long, and partly because, in this portion of it, I was merely summarizing views already expounded at greater length by Burt in *The Factors of the Mind* (pp. 31–47).[1] May we hope that he himself will deal more fully with the questions raised?

Dr. Richardson, like Dr. Eysenck and several other writers, cites Newton's *Principia* as a model; and on this point I had, in a previous contribution (this *Journal*, V, p. 208), quoted and briefly endorsed Burt's remark deprecating the unfortunate way in which "earlier psychologists—Hume, Adam Smith, Bentham, and James Mill—all, explicitly or implicitly, took Newton's hypothetico-deductive method for their model." Obviously, however, I should have offered evidence for this view. There are several reasons why Newton's method seems inappropriate in psychology.

(1) It is generally agreed that in the *Principia* Newton deliberately decided on a hypothetico-deductive method of exposition, because he thought that would be the most convincing way to

[1] I am indebted both to Professor Burt and to Dr. Richardson for reading my note in draft, and for their written comments and corrections: Professor Burt has also allowed me to make use of his lecture-notes for references and illustrations. But for the applications made in reference to the present issue I alone am responsible. Incidentally, I am puzzled by the importance Burt attaches to 'Bacon's method': Jevons, de Morgan, and Welton all assure us that "it has *not* been followed by any of the great masters of science" (*Principles of Science*, p. 507).

present his theory to his own contemporaries : instead of " the new inductive method," he adopted the " classical geometrical method," which started with definitions and axioms (or " laws "), and then deduced a succession of theorems by a train of demonstrative reasoning. This would be the only orthodox procedure for the Cartesians, who would form his main critics.

(2) The phenomena with which Newton was concerned in the *Principia* were far simpler than those dealt with by the student of individual psychology. The number of variables involved—a few large bodies each possessing mass, and moving with the passage of time in three-dimensional space—was sufficiently small for a comparatively exact analysis to be made by relatively simple mathematical procedures. The most rigorous behaviourist would never attempt to " deduce " the orbit of a golfer hunting for his ball from three hypothetical laws of motion and one universal principle in the way Newton attempted to " deduce " the orbit of the moon.

(3) What Dr. Richardson calls Newton's " long lasting theory about the motion of the planets " is not a ' hypothesis,' in the sense in which I myself was using that highly ambiguous term. I was thinking rather of what Dr. Eysenck had called a ' hypothesis,' namely (in W. E. Johnson's phrase), " the initial formulation of a general proposition which it is proposed to establish." In the terminology adopted by Kneale (*Probability and Induction*, 1949, pp. 92f.), Newton's theory of universal gravitation is a ' transcendent hypothesis ' to be established by ' *secondary* induction ' : we accept it chiefly because it co-ordinates a host of empirical generalizations by means of a simple and fruitful set of postulates. But what I had in mind was an ' empirical generalization ' of comparatively narrow range, such as might be established by ' *primary* induction ' as the result of a single factorial research.

(4) Nevertheless, it is not quite right to suppose that Newton considered it unimportant to disprove alternatives. Some had already been disproved by his predecessors. For example, Galileo's investigation of the problem of falling bodies proceeds in this way : indeed it provides the first instance of what has been called the hypothetico-disjunctive method—that combination of logical reasoning and planned experiment which is so distinctive of modern inductive science. Everyone knew that a ball rolling downhill would travel with increasing velocity ; but Galileo's problem, like that of the quantitative psychologist of to-day, was to make this and other common observations more precise by expressing them in numerical form. There were two simple hypotheses that called for consideration : (i) that the velocity is proportional to *distance* traversed ; (ii) that it is proportional to the *time* taken. He deduced testable corollaries from each, and by ingenious experiments refuted the first hypothesis and verified the second. From this he went on to disprove a still older hypothesis, namely, that force is necessary to *maintain* motion (as Aristotle's followers had assumed), and so to show that force is necessary only to *change* motion. His resolution of the motion of a projectile into two component motions at right angles—one vertical and the other horizontal—in order to test his hypothesis that the path was a parabola—is the earliest instance of factor analysis.[1] Kepler certainly did not neglect the refutation of alternative hypotheses : we are told that he tried nineteen different curves to fit Brahe's observations on the motions of Mars, before he accepted the ellipse. Newton's aim was to find a ' transcendent hypothesis ' to co-ordinate these ' empirical generalizations.' Yet he himself formally disproves the chief hypothesis which in those days might have been regarded as a serious rival to his own, namely, Descartes' theory of vortices ; and a considerable part of Book II of the *Principia* is devoted to showing that this theory is inconsistent with Kepler's laws.

Moreover, Newton's own statements about the nature of inductive proof appear to recognize the same two points that were emphasized in my own discussion, namely, the desirability of confirming the hypothesis to be established and of confuting the alternative hypotheses. For example, in his letter to Oldenburg (Secretary to the Royal Society) about " the proper method for enquiring into the properties of things," he writes that, in his opinion, the appropriate way to establish a theory is " not by deducing it only from a confutation of contrary suppositions, but also by deriving it from experiments concluding positively and directly " (Macclesfield, *Collection*, II, p. 326).

Since Newton is so often cited in discussions on scientific methodology, it is perhaps worth while glancing at the procedure which he himself actually adopted in one of the more important of his own experimental researches (*Phil. Trans.*, 1672, LXXX, p. 3075). The current theory of light (as expounded, for instance, in Kepler's *Dioptrice*) supposed that pure light was essentially white, and that colour was a " qualification of light " derived from the process of refraction. To test this view Newton began by making a small circular hole in the window-shutter of a darkened chamber, permitting the entrance of a narrow beam of sunlight. This produced a bright circular patch on the opposite wall. He then placed a prism near the entrance. " According to the received laws of refraction," he says, one would have expected the patch still to be " circular," though shifted by refraction from its former position and possibly tinged with colour. Actually the prism produced " a coloured spectrum " of " an oblong form." He then considers whether certain supplementary hypotheses would suffice to explain these peculiarities, e.g., " whether the rays, after trajection

[1] No doubt on many of these points casual anticipations could be found in the work of da Vinci, Stevinus, and others.

through the prism, did not move in curved lines." By varying the size of the hole and the thickness of the glass, he was able to eliminate each one. " The gradual removal of these suspicions," he says, " led me to the *experimentum crucis*." This consisted in placing a second board, containing another small aperture, beyond the prism ; and then " turning the prism slowly about its axis." He also subjected the separated rays to a second refraction, showing that each preserved its colour, and that together they could be recombined to give white light. The various results conclusively disproved the corollaries from the accepted theory, and confirmed the corollaries deducible from his own hypothesis, namely, that " light itself is a heterogeneous mixture of differently refrangible rays " and that " the same degree of refrangibility ever belongs to the same colour." It is noteworthy that Hooke criticized Newton's logic on the ground that the explanation reached was still " not the only hypothesis. . . . The same phenomenon will be solved by my hypothesis as well as his." In particular, as is now fully realized, Newton in his discussion about the phenomena of light (e.g., those of polarization) failed to consider that the undulatory hypothesis might take two forms : the waves need not be longitudinal (as Huyghens and Newton assumed) ; they might also be transverse (as Hooke tentatively suggested).

If Newton failed, it is scarcely surprising that Dr. Richardson should ask : " how can anybody be *certain* that he has thought of *all the possible alternatives ?* " And Burt, in the chapter that I quoted, also observes that " the difficulty is to be sure that our enumeration of the possible competitors is exhaustive." However, he suggests certain plausible means by which this difficulty may be largely overcome. To begin with, he reminds us that in an inductive enquiry we can never hope to " be certain." We can only strive to reach the most probable conclusion. For this purpose we must begin by " considering the *antecedent* probabilities of the competing hypotheses." These may differ so widely as to justify us in abandoning any attempt to examine " *all* the possible alternatives," and in confining ourselves to those that have a high degree of initial probability. Moreover, if a hypothesis (like Dr. Richardson's theological alternative) can be neither confirmed nor confuted, then it is not a hypothesis that the empirical scientist is called upon to consider. Best of all we may be able to " plan the research so that only two alternatives are left open, namely, the hypothesis to be proved and what is commonly called the null hypothesis."

Burt illustrates[1] his argument by deriving (from the familiar rule for compound probabilities) the following formula :

$$P_a = \frac{1}{1 + \frac{p_n q_n}{p_a q_a}}$$

where a denotes the hypothesis to be established, n its contradictory (e.g., the null hypothesis), p the initial probabilities, q the likelihoods after the experiments have been made, and P_a the final or posterior probability. We may assume that neither p_n nor p_a is zero, and (in many cases at any rate) that they do not differ greatly from each other : otherwise we should hardly have embarked on the research. But, except perhaps in Mendelian genetics, they can seldom be determined precisely in any psychological enquiry. If the contradictory hypothesis had the form of a universal proposition (e.g., there are *no* group factors underlying psychotic conditions), it could conceivably be disproved outright by a single observation ; q_n would be zero, and $P_a = 1$. But this is exceptional. Again, if our experiment can be planned so as to verify a corollary that can be deduced logically from hypothesis a, then q_a will be 1 ; and P_a will depend mainly on the size of q_n. This is the situation that we aim at when we try to establish hypothesis a by disproving Fisher's ' null hypothesis ' (i.e., the hypothesis that the results obtained are attributable to chance). It is obvious that the final probability may in some measure be improved either by seeking merely to confirm a (making q_a large) or by seeking merely to disprove n (making q_n small) : but the improvement will then as a rule be small, and in general we are not likely to raise P_a much above half unless we seek to do both.

In the particular problem with which Dr. Eysenck and I were concerned, the hypothesis to be proved was that " all forms of psychoses are merely different degrees of one and the same underlying condition, i.e., the differences are quantitative, not qualitative." Here I fancy most psychiatrists would hold that p_n is much higher than p_a. In any case I do not wish to deny that the hypothesis demands that we ought first to seek (as Dr. Eysenck has done) evidence supporting a dominant general factor (thus raising q_a). But that by itself is not enough. In my view it should also be possible to plan a factorial research so that, if there were after all qualitatively different types of psychosis, then we should find corresponding group factors that are statistically significant ; while if there were not, then the residuals, even with large samples, would be non-significant : (and in this latter case we should have diminished q_n). I would also add that, to carry complete conviction, the research

[1] In a written note he explains that he regards his formula as illustrative only : " Kneale has argued that the ' probability ' of inductions is not the same as the ' probability of chances,' and others have declared that, even if they were, we could not compute them exactly : nevertheless, I hold that the formula may illustrate the way in which what Kneale calls the ' degrees of acceptability ' may be lowered or raised : (cf. *Probability and Induction*, pp. 211f., 250f.)."

should test the alternative hypotheses, not only by factorial procedures, but other methods as well ; for, even if a statistical enquiry fails to find qualitative differences, the psychiatrist will still wonder whether clinical, biochemical, or therapeutic methods might not have been successful. In short, as Kneale puts it, " the scientist's aim is to make a theory which fits *all* the facts, and at the same time *has no rival* " (*op. cit.*, p. 253 ; his italics).

I am not wholly satisfied with Dr. Richardson's ingenious way of getting rid of rival theories, namely, by a kind of mutual extermination *a priori*. It would certainly provide a quick solution to the controversy between Dr. Eysenck and myself. Following Dr. Richardson I could argue that " the theory prevalent in the U.S.A." is that all correlations between mental traits are " caused entirely " by group factors ; and that another theory, prevalent in monistic circles, is that they are " caused entirely " by a single general factor. Then I could ask : " do they not refute each other at least so far as the adverb ' entirely ' is concerned ? " If so, what is left would be a general-*plus*-group factor theory, such as I have supported. In this manner the problem would be disposed of straight away, without the need for any empirical evidence whatever ! But would Dr. Richardson be convinced ?

In his own interesting investigation (if I understand his reasoning correctly) he has taken the three hypotheses which seem most plausible *a priori* : then, on deducing their consequences, he has shown that the results of the third seem to agree with facts, whereas those of the first and the second issued in absurdities. W. L. G.

Editorial Note.—The correspondence on this subject in the previous issue has prompted several letters, asking for further references on the issues raised. On the fundamental question of scientific method, there can be little to add to the full discussions in Russell's *Human Knowledge* (esp. Pt. VI, ' Postulates of Scientific Inference ') and the works he there quotes—by Keynes, Nicod, and Broad. Here it is scarcely possible to reply in detail to the questions put by the various correspondents or by the reviewer himself in his opening paragraph. Since, however, he has referred to my own opinions, it may save some misunderstanding if I briefly indicate one or two minor points on which I should disagree.

I cannot help feeling that, in restricting himself chiefly to physicists for illustrations of the logical procedure most appropriate to psychology, he has laid himself open to much the same kind of criticism that he urged against Dr. Eysenck. Jeffreys, in the passage cited by Dr. Richardson, has shown that the methodology of the earlier physicists was by no means so faultless as their psychological imitators have supposed ; and certainly it is hardly the methodology most suitable for the psychologist. For early examples of the kind of reasoning appropriate to psychological experiments, I should prefer to go to biological writers, like Harvey or Cæsalpinus, rather than to physicists like Newton or Galileo.

In the reply to Dr. Richardson it is alleged that " the first instance of the hypothetico-disjunctive method " is to be found, not in Newton (with which I should agree), but in Galileo : much the same statement is to be found in popular introductions and in books by continental psychologists.[1] But as a matter of history, this seems incorrect. So far as European writers are concerned, the revolution in methodology was not so much the dramatic achievement of any single genius ; it was the outcome of a new tradition that had been slowly gathering force, in the face of vigorous criticism, at the University of Padua—a tradition which is traceable back to Buridan of Paris, and so to Buridan's English master, William of Ockam, whose ' razor ' implies that hypotheses are more often to be rejected than accepted.[2] The change itself is in part an outcome of the prolonged disputes between Franciscans and Dominicans, which brought methodological issues to the fore.

As one of the earliest applications of the ' method of exclusion,' I should be inclined to cite, not Galileo's *Demonstrazione Matematiche* (1638), but Harvey's *De Motus Cordis et Sanguinis* (1628).[3] The opening sentence of his introduction declares that it will be imperative to begin by stating the

[1] Bertrand Russell, for example, writes : " Scientific method, as we understand it, comes into the world full-fledged with Galileo " (*The Scientific Outlook*, p. 22 : cf. also Lewin, K., *A Dynamic Theory of Personality*, ch. I, ' The Conflict between Aristotelian and Galilean Modes of Thought in Contemporary Psychology '). Just as English writers, from Hume onward, have urged that empirical psychology should take Newton for its model, so continental writers have put forward Galileo. Lewin is of course fully justified in his censures on Aristotle's treatment of *physical* problems ; but that has far less relevance for psychology than psychologists are apt to suppose. In regard to the discussion of *biological* problems (of which his *De Anima* was a part) I prefer Darwin's verdict : " Linnæus and Cuvier may have been my gods, but they were mere schoolboys compared to old Aristotle " (*Life and Letters*, III, p. 252).

[2] Cf. Duhem, P., Études sur Leonard de Vinci (1906–13) and Randall, J. H., ' The Development of Scientific Method in the School of Padua,' *Hist. of Ideas*, I, 1940.

[3] ' Le plus beau livre de la physiologie,' as Flourens called it. Harvey and Galileo were at Padua together ; and that University had already produced several revolutionary writers in the biological field, such as Vesalius, Cæsalpinus, Colombo, and Fabricius.

chief alternative suppositions " in order that the false [e.g., Galen's hypothesis of ebb and flow] may be disproved, and the true [viz., Harvey's alternative hypothesis of circulation] may be confirmed, by accurate observation and multiplied experience " ; and his unexpected introduction of quantitative proofs into what is a problem in biology rather than physics is a particularly instructive feature of his reasoning.

As regards the explicit theory of inductive methodology, there is ample evidence (in spite of the reviewer's objections) that the change in standpoint was mainly due to Bacon, who in later years was Harvey's patient. Bacon is the earliest writer who definitely insists that " the first duty of induction is rejection or exclusion " (*Novum Organon*, 1620, II, Aph. 16 ; cf. Sigwart, *Logic*, II, p. 296). Jevons, like most nineteenth century writers, greatly underrated Bacon's early influence. As Reid observes, " the best models of inductive reasoning, such as the 3rd book of the *Principia* and the *Opticks* of Newton, were drawn from Bacon's rules " [1] : Newton's very phrase ' *experimentum crucis* ' is borrowed from the *Novum Organon* (II, Aph. 36). Hume, in the introduction to his *Treatise*, describes the application of the ' experimental method of reasoning,' first to simpler sciences, and then ' at the distance of a century ' to ' the science of man,' as having originated in the new principles laid down by ' my Lord Bacon ' ; and he himself constantly uses the disjunctive type of reasoning.[2] The psychologist will recall that Bacon announced that his inductive method was to be applied ' non tantum ad naturales, sed ad omnes scientias,' expressly including psychology and social or political theory (I, Aph. 127). Finally, nearly all the eminent scientists associated with the early work of the Royal Society—Boyle, Hooke, Newton, Wren, and Wilkins, together with Petty and Grant, the founders of ' Politicall Arithmetic '—acknowledge the lead of the *Artium Instaurator* [3] : and a contemporary ballad describes how

> The Prime Virtuosi have undertaken
> Through all the experiments to run
> Of that learned Man, Sir Francis Bacon,
> Shewing what can, what can't, be done.

For another hundred years a heated conflict dragged on between the English inductive school, which claimed Bacon as its founder, and the French deductive school, deriving from Descartes. But by the middle of the 18th century the inductive method had won ; and Bacon is restored to his pedestal by D'Alembert and the *Encyclopédistes*.

Yet once again the formulation of the methods proper to the inductive sciences cannot be credited to any single thinker. It can only be educed from an historical and critical review of such methods as they have actually been worked out and practised by scientists themselves. Nor is there as yet any final or satisfactory formulation to which the novice can be referred. As Russell has remarked, " the inductive method is the glory of science, but the scandal of philosophy."

In conclusion, may I add that I readily agree with Dr. Richardson that there is still something that the psychologist may learn from Newton's methodology in dealing with the problems of motion, provided we reinterpret his procedure in the light of the later developments to which Dr. Richardson himself alludes ? It has, I think, become clear that the so-called ' laws ' are not so much ' hypotheses ' to be established by inductive or deductive procedures, but simply ' principles ' stating how the scientist proposes to deal with the phenomena in question : they express, not the conclusions of his inferences, but the prerequisites of his inferences. When, in dealing with optical phenomena one man chooses as his fundamental concept that of the light-ray, while another chooses that of the light-wave, neither is propounding a generalization which we can show to be either true, or false, or merely probable : each is announcing a working policy he has determined to adopt.[4] The same is true when Riemann and later Einstein, discarding ' action at a distance,' decide to substitute a ' field theory.'

A recognition of this methodological distinction might, I fancy, go far to ease some of the current disputes in psychology. In the statistical field there are obvious analogues in the use of factor analysis. The fundamental principles preferred by advocates of this or that factorial method are not to be regarded as generalizations inferred from observational data, any more than the principle of the syllogism is a generalization inferred from empirical phenomena ; they indicate the kind of

[1] *Works* (1785) : p. 200 in Hamilton edition. If a contemporary testimony is desired, I may cite Professor Broad's considered estimate : " In his analysis of inductive arguments Bacon was, so far as I know, breaking new ground ; and all later discussion has followed on his lines " (*Selected Essays : Ethics and the History of Philosophy*, 1952, pp. 141–2).

[2] E.g., *Treatise on Human Nature*, I, iv, 5 (' On the Immateriality of the Soul ') : " The absurdity of the last two suppositions sufficiently proves the veracity of the first ; nor is there any fourth opinion."

[3] Bacon, so named, figures in the frontispiece designed by Evelyn for Sprat's *History of the Royal Society* (1667). The ballad (British Museum) is attributed to J. Glanvill.

[4] This view has been urged (in somewhat paradoxical terms) by F. P. Ramsey and M. Schlick : cf. also Popper, K. R., *Logik der Forschung*.

simplification each proposes to adopt in constructing a workable map or model of the phenomena he is trying to describe or understand. The choice is rather like that of the maker of geographical maps or globes : according to his purpose he may seek to locate positions on the earth in terms of a three-dimensional sphere, of a curved space of two dimensions, or of an orthographic, stereographic, or Mercator projection. And it would be plainly misleading to treat such devices as general propositions that must be formally proved or disproved by some process of inductive reasoning.

Having decided on the general line of approach and defined the working concepts it entails, we then can proceed to set up and test alternative ' hypotheses ' (in the narrower sense of the term) ; and here I think I am in almost complete agreement with what has been said above about the merits of the ' hypothetico-disjunctive ' procedure. For indications as to how precisely it should be employed in any particular research, the reader may be safely referred to Fisher's imaginary ' psychophysical experiment ' as described at the beginning of *The Design of Experiments* (1937, pp. 13–29), or, for more detailed working instructions, to the section dealing with ' Statistical Hypotheses and their Verification ' in the recent volume on *Statistical Inference*, by H. M. Walker and J. Lev (1953, pp. 5f.).

C. B.

THE RELATIVE MERITS OF RANKS AND PAIRED COMPARISONS

To the Editor, British Journal of Statistical Psychology

Sir,—Recent articles in this *Journal* (Durbin, IV, pp. 85–90 ; Whitfield, VI, pp. 35–40) suggest that the ranking-methods which were introduced into psychology 50 years ago by Spearman and which were so severely criticized by contemporary psychologists are regaining their old popularity. They are unquestionably attractive to the practical worker ; but he would like to be assured that the former criticisms have been adequately met. Titchener and his followers (e.g., Myers and Valentine in this country) maintained that paired comparison was far superior as an experimental technique, since it enabled the investigator to discover preferences, etc., that were psychologically compatible, though logically inconsistent : Brown, in *The Essentials of Mental Measurement*, wrote still more emphatically : " the use of ranks as a measure of a characteristic is inadmissible . . . Some form of frequency distribution must be assumed ; the method of ranks assumes it to be a rectangle. The assumption could hardly be more improbable."

Dr. Whitfield and others, who apparently favour the method, appear to disregard these various objections. Moreover, Dr. Whitfield's interesting article in the last issue of this *Journal* raises several additional questions. He shows that the measurements derived from summing m rankings tend, when m is large, to produce a normal distribution. Does not this imply that, if (as Dr. Brown observes) " some form of frequency distribution must be assumed," it would be safer to assume that the distributions are normal ? And how are we to calculate the correlation between such a sum and an independent criterion which itself may be given in the form of ranks ? Should we re-rank the sums, thus substituting ordinal numbers for cardinal, or should we rather convert the criterion ranks to normal deviates by using Hull's table, as described in so many text-books of psychological statistics ?

Dr. Whitfield's discussion of his central problem still leaves many steps somewhat obscure. He mentions two alternative procedures available for testing the statistical significance of a measurement derived from the summation of m ranks, first, that given by Professor Kendall, and secondly, one given by himself and Professor Burt. He somewhat curtly declares that " Kendall's treatment of the problem of m ranks is inappropriate." But he offers no reasons for rejecting this solution or for preferring the second. He quotes Burt as having derived a formula for the frequencies to be expected by chance, but again no proof is given. Since the new solution is of special interest to practical psychologists, could either Professor Burt or Dr. Whitfield insert the missing steps, and could they also explain why the one solution is not appropriate and how the other is deduced ?—Yours, etc., R. E. Morris.

Reply.—1. The main objections to the technique of ranking were raised not so much by psychologists as by statisticians, notably by Karl Pearson [1] ; and it is the statisticians, more particularly Professor Kendall, who have restored it to respectability. Brown was unquestionably right in stating that ranks do not form a direct ' quantitative measure ' of the quality to which they refer. Ranking is rather to be regarded as a special type of ' manifold classification '—one in which the classes can be arranged in definite order. The use of ties will permit more than one individual to be placed in the same class and so receive the same rank ; and, if the appropriate number of ties is forced on the ranker, then his ranks *can* (if we wish) be made to yield any type of distribution, including the normal. But, as Galton pointed out, for statistical purposes an individual's rank-number

[1] Cf. Pearson, K., *Further Methods of Determining Correlation* (Drapers' Company Research Memoirs : Biometric Series, IV, 1907). Brown's criticisms, as he himself observes, are largely a summary of Pearson's. Incidentally, the method of ranks was not " introduced into psychology by Professor Spearman," but by Galton, and popularized by Galton's younger collaborator, J. M. Cattell : (see Guilford, *loc. cit. inf.*, pp. 244f., and refs.).

(or rather his rank-number *minus* one) can best be treated, not as an ordinal number, but as a cardinal number : it states the number of individuals in the classes prior to his.[1] Although from a practical standpoint the distinction may seem somewhat academic, it is nevertheless important for the rigorous deduction of certain formulæ. Thus, with this interpretation of rank-numbers, no assumption need be made about the form of distribution (as Brown erroneously supposed). If so, it follows that, in addition to ease and speed of working, the use of ranks has a special theoretical advantage of its own : it provides a convenient *non-parametric procedure*, i.e., a procedure which can be used when we desire to avoid any specific assumption either about the distribution of the variables or about the scale of measurement. In particular, the analysis of variance takes a very simple form when the data are expressed in terms of ranks.[2]

2. Let me now deal with the most fundamental of the difficulties raised by Mr. Morris. In his last paragraph he enquires why the ' solution ' derived by Dr. Whitfield and myself was preferred to that proposed by Professor Kendall. The explanation is that the two solutions are really solutions to two different problems, namely, the treatment of summed ranks (*a*) for individuals and (*b*) for groups.

The difference can be explained more precisely if we take a concrete case. Let us suppose that an education authority has adopted the ' quota system ' of allocating children to a grammar school : as a result, School A is allowed (let us say) two pupils, and the teachers are required to name the two brightest individuals out of a batch of eight who happen to have reached the appropriate age. The headmaster accordingly suggests that the four teachers who know all the pupils should each rank them in order of merit, and that the two boys with the highest average rank should be selected.

TABLE I. TEACHERS' RANKS FOR EIGHT BOYS

Boy	Examination	Teachers i	ii	iii	iv	Total for Teachers	Average
T.S.	1	2	2	1	3	8	2
B.R.	2	1	1	3	2	7	$1\frac{3}{4}$
O.F.	3	4	6	8	4	22	$5\frac{1}{2}$
R.L.	4	5	4	7	5	21	$5\frac{1}{4}$
E.B.	5	8	5	6	1	20	5
M.W.	6	7	3	5	8	23	$5\frac{3}{4}$
G.A.	7	6	7	2	7	22	$5\frac{1}{2}$
J.H.	8	3	8	4	6	21	$5\frac{1}{4}$
Sum of Squares of Deviation	(*a*) about teachers' mean	42	42	42	42	300	$18\frac{3}{4}$
	(*b*) about boys' mean	$18\frac{3}{4}$	$19\frac{3}{4}$	$27\frac{1}{4}$	$27\frac{1}{4}$	—	—
		93					

Table I shows the orders actually given by the teachers in a case of this sort : the total and average rank for each boy are shown on the right. If we could trust the rankings, then B.R. and T.S. would seem to be the brightest of the eight. It is the statistical investigator's task to determine whether the results *can* be trusted.

Now the theoretical psychologist is interested in the trustworthiness of such rankings obtained in relation to the group as a whole ; the practical teacher is interested only in the trustworthiness of the assessments for the two top boys. The questions are plainly distinct.

(*a*) Let us consider first what is the validity of the rankings *for the whole group*. To determine this, the natural approach will be to carry out a factorial analysis : in that way we can ascertain whether there is any general component common to all four rankings, and, if so, whether it is statistically significant. The correlations between the rankings in Table I, together with the saturations deduced from them, are shown in Table II.

[1] The term ' grade ' (a word introduced by Galton) was originally used to designate position as defined by the cardinal number, and distinguished from ' rank,' which is reserved for the ordinal number. Thus Brown, following Pearson, states that " the *grade* of an individual is measured by the number of individuals above him."

[2] For a fuller discussion the reader may refer to Friedman, M., ' The Use of Ranks to avoid the Assumption of Normality,' *J. Amer. Statist. Assoc.*, XXXII, 1937, pp. 675–701, Kruskal, W. H., and Wallis, W. A., ' The Use of Ranks in One-criterion Variance Analysis,' *ibid.*, XLVII, 1952, pp. 583–621. Cf. also Friedman, M., ' A Comparison of Alternative Tests of Significance for the Problem of *m* Rankings,' *Ann. Math. Stats.*, XI, 1940, pp. 86–92.

TABLE II. CORRELATIONS BETWEEN TEACHERS' RANKINGS

Teachers	i	ii	iii	iv	Total
i	1·0000	·3333	·4048	·2619	2·0000
ii	·3333	1·0000	·2381	·3810	1·9524
iii	·4048	·2381	1·0000	− ·0477	1·5952
iv	·2619	·3810	− ·0477	1·0000	1·5952
Total ..	2·0000	1·9524	1·5952	1·5952	$7{\cdot}1428 = 2{\cdot}6726^2$
Saturations ..	·7483	·7305	·5969	·5969	2·6726

To test the significance of the general factor we may follow the procedure outlined in *Factors of the Mind* (ch. X). The average of the *observed* correlations (i.e., the average of the coefficients in Table I when the self-correlations are excluded) is given (*op. cit.*, p. 275) by the equation

$$\bar{r} = r_{\text{int}} = \frac{1}{m\,(m-1)}\,\Sigma r_{ii'} \qquad (i \neq i') \quad \ldots \text{(i)}$$

$$= \frac{1}{12} \times 3{\cdot}1430 = {\cdot}2619.$$

In order to decide whether or not this figure is statistically significant, it is convenient first to calculate a quantity which I denoted by the symbol η^2. This is given (pp. 275–6) by the equations

$$r = \frac{m\eta^2 - 1}{m-1}, \text{ or}$$

$$\eta^2 = \frac{(m-1)\,\bar{r} + 1}{m}$$

$$= \frac{1}{m^2}\,\Sigma r_{ii'} \qquad (i, i' = 1, 2, \ldots m) \quad \ldots \text{(ii)}$$

$$= \frac{1}{16} \times 7{\cdot}1428 = {\cdot}4464,$$

where $\Sigma r_{ii'}$ includes the m self-correlations (all taken = 1) as well as the observed inter-correlations. It follows that $\eta^2 = \frac{(\Sigma r_{ij})^2}{m^2}$ (p. 276). This quantity is the square of the correlation ratio, and could, if we preferred, be calculated by the more usual formula direct from the original data. We have in fact (p. 274)

$$\eta^2 = \frac{Q_m}{Q_t} \qquad \ldots \text{(iii)}$$

(where Q_m denotes the sum of the squares of the deviations about the means, each squared deviation being taken m times, and Q_t denotes the sum of deviations of each individual mark or rank about the mean for each teacher)

$$= \frac{4 \times 18\frac{3}{4}}{4 \times 42} = \frac{75}{168} = {\cdot}4464, \text{ as before.}$$

Alternatively, if we write Q_r for the sums of the squares of the residual deviations (93), so that Q_m and Q_r are the ' between groups ' and ' within groups ' square-sums, then

$$Q_t = Q_m + Q_r = 75 + 93 = 168\,;$$

and we thus have an alternative method for determining the denominator.

To test the significance of the factor (p. 274) we can now use the procedure adopted for testing the significance of a correlation ratio. Accordingly, we shall take

$$F = \frac{V_m}{V_r} = \frac{Q_m \div (n)_f}{Q_r \div (nm)_f} = \frac{Q_m \div (n-1)}{Q_r \div (n-1)\,(m-1)}$$

(where V_m and V_r denote the variance ' between groups ' and ' within groups ' respectively, and $(n)_f$ and $(nm)_f$ denote the appropriate degrees of freedom)

$$= (m-1)\,\frac{Q_m}{Q_r} = (m-1)\,\frac{n^2}{1-n^2} \qquad \ldots \text{(iv)}$$

$$= 3 \times \frac{75}{93} = 3 \times {\cdot}8064 = 2{\cdot}4192.$$

On referring to Snedecor's table for the distribution of F, we find that this figure is somewhat below the value given for F at the 5 per cent. level. We are therefore forced to conclude that the factor is *not* statistically significant.

With a larger table we could use a chi-squared test. $m(n-1)\eta^2$ may be regarded as the sum of the squares of a number of standardized deviations ; and, as m increases, the distribution of this value tends to that of chi squared. Let us therefore calculate

$$\chi^2 = m(n-1)\eta^2 = \frac{n-1}{m}\Sigma r_{ii'} \quad (i, i' = 1, 2, \ldots m) \ldots \text{(v)}$$

$$= 1{\cdot}250.$$

Entering the chi-squared table with $n-1$ degrees of freedom, we find once again that this value is definitely below that required for the 5 per cent. level.

There are two convenient working formulæ that often help to shorten the calculations. To avoid calculating the deviations of the totals (which often introduce fractions), it is quicker to square the sums as they stand, and then correct for the mean. If s_i denotes the total of the rank allotted to the ith boy, then :

$$\eta^2 = \frac{1}{n-1}\left\{\frac{12\Sigma s_i^2}{m^2n(n+1)} - 3(n+1)\right\} \quad \ldots \text{(vi)}$$

and

$$r = \frac{1}{(m-1)(n-1)}\left\{\frac{12\Sigma s_i^2}{mn(n+1)} - (3m+1)(n+1) - 2\right\} \quad \ldots \text{(vii)}$$

The general procedure outlined above is very similar to that suggested by Kendall [1] in dealing with " the case of m rankings." He proposes to calculate what he terms a ' coefficient of concordance,' expressed by the equation

$$W = \frac{\text{Sum of squares of the deviations of the total ranks about their means}}{\text{The maximum value possible for such a sum}}.$$

In my book (*loc. cit.*, p. 275) I pointed out that Q_t, defined as above, is also the maximum value of Q_m. It follows that Kendall's coefficient

$$W = \frac{Q_m}{Q_t} = \eta^2. \quad \ldots \text{(viii)}$$

He also gives eqn. ii and iv (with W written for η^2) ; notes that r is equivalent to the intra-class-correlation ; and proposes to use Fisher's z-test to test the significance of the variance-ratios. This, of course, is equivalent to the F-test. It would seem therefore that the two methods of dealing with this problem, reached from two somewhat different standpoints, are really identical.

(*b*) The issue which confronts those who have to decide whether or not this pupil or that deserves a scholarship is quite distinct. They are concerned, not with the statistical significance attaching to the figures for the group as a whole, but with the statistical significance of the figures for some *particular individual.* In the present example we are required, if possible, to pick out two boys, and two boys only, for a possible scholarship award. According to the summed ranks, B.R. is the best and T.S. nearly as good ; but, if we are to justify this choice, we must be able to show that the difference between the figure for T.S. and the average for the whole group is statistically significant. Whether or not the detailed ratings of the other six boys are statistically trustworthy no longer matters.

Consider, then, the total mark obtained by the boy T.S. (who is second in order of merit), namely, 8. This is a deviation of -10 from the average mark, namely, 18. Could such a deviation have arisen by chance ?

To decide this question, let us suppose that the teachers have allotted their ranks to T.S. entirely at random. This can best be done by adopting some mechanical method of randomization. We may accordingly imagine that each teacher uses an octagonal teetotum, with its sides labelled from 1 to 8 ; and, having spun it once, takes whatever number comes uppermost to represent T.S.'s rank. The four random ranks are then summed. Is it likely that a value so high as 8 would thus be secured by sheer chance ? In order to obtain an answer, we require first of all to determine the frequency-distributions of such sums when obtained by purely haphazard assignment.

[1] Kendall, M. G., *Advanced Theory of Statistics*, I (1943), p. 411 ; cf. *Rank Correlation Methods* (1948), ch. VI, and refs. The procedure in my book was intended to be of general application. Kendall has dealt specifically with data in the form of ranks ; and, in addition to the more systematic mathematical investigation of this problem, has shown that the procedure may also be used in the case where the ranks are tied.

Mr. Morris enquires how I derived the formula for such frequencies. The problem is an ancient one in the theory of games of chance (see Todhunter, *History of Probability*, 1865, art. 987) : it is simply a generalization of the more familiar problem of tossing a coin whose two sides yield two alternative values only. As is well known, when there are only two alternatives, the distribution is given by the *binomial* theorem ; when there are more than two, the same reasoning shows that the distribution will be given by the *multinomial* theorem.

Let the number of alternatives (i.e., the number of sides to the teetotum) be n ; let the number of teetotums spun (i.e., the number of teachers judging) be m ; and let s be the sum of the m ' ranks ' obtained as a result of the m spins.

Since any one of the n faces may turn uppermost on each of the m teetotums, the number of different ways in which they may fall after spinning will be n^m. To determine the number of ways in which the sum may be obtained, the simplest procedure is to use the device of a " generating function." [1] We find that the number required will be equal to the coefficient of t^s in the expansion of

$$\begin{aligned}&(t^1 + t^2 + t^3 + \ldots + t^n)^m\\ &= t^m(1 + t^1 + t^2 + \ldots + t^{n-1})^m\\ &= t^m(1 - t^n)^m\,(1 - t)^{-m}. \qquad \ldots \text{(ix)}\end{aligned}$$

The coefficient of t^s in this expansion will be the coefficient of t^{s-m} in the expansion of $(1 - t^n)^m\,(1 - t)^{-m}$.

Now $$(1 - t^n)^m = 1 - mt^n + \frac{m(m-1)}{1.2}\,t^{2n} - \frac{m(m-1)\,(m-2)}{1.2.3}\,t^{3n} + \ldots;$$

and $$(1 - t)^{-m} = 1 + mt + \frac{m(m+1)}{1.2}\,t^2 + \frac{m(m+1)\,(m+2)}{1.2.3.}\,t^3 + \ldots$$

Multiply the two series together and pick out the coefficient of t^{s-m} in the product. This is easily found to be

$$\begin{aligned}&\frac{m\,(m+1)\ldots(s-1)}{(s-m)\,!} - m\,\frac{m\,(m+1)\ldots(s-n-1)}{(s-n-m)\,!}\\ &+ \frac{m(m-1)}{1.2}\cdot\frac{m(m+1)\ldots(s-2n-1)}{(s-2n-m)\,!} - \ldots, \qquad \ldots \text{(x)}\end{aligned}$$

the series being continued so long as no negative factors appear in the products. The probability of obtaining s is then obtained by dividing the series by n^m. This, I hope, answers the enquiry made, not only by Mr. Morris but also by several other correspondents who have asked how the formula in the footnote on p. 36 of Vol. VI was actually reached.

In practice, since we shall commonly be concerned with the tail-end of the distribution, there will seldom be more than two or three terms in the series. Here, for example, $s - m = 8 - 4 = 4$; and this is smaller than n. Hence we need consider only the first term of the series ; and its value can be obtained from the successive coefficients in the above expansion of $(1 - t)^{-m}$. Their values in turn can be found by taking the 1st, 2nd, . . . figures in the usual table of binomial coefficients (e.g., in *Barlow's Tables*, p. 207), starting with the 1st coefficient for $n = m - 1$. Here, therefore, they are 1, 4, 10, 20, 35. The sum is 70 ; and dividing this by $n^m = 4096$, we obtain $P = 0{\cdot}01709$ (as given in Whitfield's table).

Were the numbers large enough for us to assume that the distribution of the sum values was approximately normal, we could argue as follows. The mean of the summed ranks is $\frac{1}{2}m\,(n + 1) = 18$ (as may be verified from the last column but one in Table I) ; the variance is $\sigma^2 + \frac{1}{12}m\,(n^2 - 1) - 21$; the absolute value of the deviation (d) will be

$$|\text{ sum of ranks for boy } - \tfrac{1}{2}m\,(n-1)\,| - \tfrac{1}{2} = |\,8 - 18\,| - \tfrac{1}{2} = 9\tfrac{1}{2}$$

(where the subtraction of $\frac{1}{2}$ is the correction required in the case of a discontinuous variable). Hence $d/\sigma = 2{\cdot}07$, which gives $P = 0{\cdot}019$. Thus the attempt to discriminate the two best boys *is* statistically significant, although the attempt to discriminate the individuals in the group as a whole was *not* statistically significant. With this particular group the explanation of the divergence is simple : as the headmaster put it in discussing the results, " except for the two lads T.S. and B.R., there is little to choose between the various members of this class, since the duller pupils of that age were left in the class below."

3. Mr. Morris enquires about the procedure to be followed when we have " an independent criterion which is itself given in the form of ranks." As it happens, all the eight boys in the group described above eventually sat for the final examination for junior county scholarships ; and a mark-list giving all the candidates in order of merit was circulated to the board of examiners (of whom, as L.C.C. pyschologist, I formed one). Accordingly in Table I, I have prefixed a column showing

[1] Cf. Aitken, A. C., *Statistical Mathematics* (1939), pp. 16f.

the ranking of these boys in this later examination ; and we may regard this as an " independent criterion " of the kind he has in mind.

Let us therefore suppose that we wish (as Mr. Morris suggests) to calculate some kind of multiple correlation between (*a*) the composite result of the school teachers' judgements, calculated as an unweighted sum of their several ranks, and (*b*) the results of the final examination, expressed as an order of merit. How are we to proceed ?

Obviously, to convert the sums to ranks, or the ranks to normal deviates,[1] would lower the correlation. The simplest method is to take the unweighted sums as they stand. Let s_i (as before) denote the sum of the teachers' ranks, and c_i the criterion rank ; then the ordinary product-moment formula will yield the following equation :

$$r_m = \frac{4\Sigma a_i s_i - mn(n+1)^2}{\sqrt{\left\{\frac{1}{3}n(n^2-1)\left[4\Sigma s_i^2 - m^2 n(n+1)^2\right]\right\}}}. \qquad \ldots \text{(xi)}$$

E.g., with the data given in Table I, we have

$$r_m = \frac{4 \times 732 - 4 \times 8 \times 81}{\sqrt{\left\{\frac{1}{3} \times 8 \times 63\left[4 \times 2892 - 16 \times 8 \times 81\right]\right\}}}$$
$$= \cdot 748.$$

This formula provides a quick and useful way of reaching an approximate estimate for the multiple correlation with a criterion : (it is actually its lower limit). To find the best value we should have to weight the rankings furnished by each teacher ; and these weights would be computed from the correlations and the inter-correlations in the usual way.

To ascertain the improvement that has been effected by basing the correlation on all the m rankings simultaneously, instead of on a single ranking only, we can calculate the *average* of the m correlations that would be obtained by taking each teacher's ranking singly. There is no need to calculate the correlations explicitly. The average required can be obtained direct by means of the following formula :

$$r_{av} = \frac{3}{n-1}\left\{\frac{4\Sigma a_i s_i}{mn(n+1)} - (n+1)\right\}. \qquad \ldots \text{(xii)}$$

E.g., with the figures given above we have

$$r_{av} = \frac{3}{7}\left\{\frac{4 \times 732}{4 \times 8 \times 9} - 9\right\}$$
$$= \cdot 500.$$

In this case by taking four teachers' judgements instead of only one we have improved the ' coefficient of prediction,' as it has sometimes been called, by nearly 50 per cent. We can check the accuracy of the foregoing formula by calculating the four correlations separately, and then finding their average in the usual way. We obtain

$$\tfrac{1}{4}(\cdot 524 + \cdot 762 + \cdot 071 + \cdot 643) = \cdot 500.$$

Since the distribution of the average rank correlation approaches a normal distribution more quickly than that of a single rank correlation, we may, as a rough test, take the standard error to be

$$= \sqrt{\frac{1}{m(n-1)}} = \frac{1}{\sqrt{28}} = 0{\cdot}189.$$

The value found is therefore statistically significant.

On re-writing the equation for the multiple correlation (xi) in a form analogous to that of the equation for the average correlation (xii) it will be seen that the numerator is the same in both cases. The denominator differs by the substitution of

$$\sqrt{\left\{\frac{1}{m}\left[4\Sigma s^2 - m^2 n\ (n+1)^2\right]\right\}} \text{ for the factor } \sqrt{\left\{\frac{1}{3}mn(n^2-1)\right\}}.$$

On comparing this with eqn. vi and viii we note that the increase of the multiple correlation over the average correlation depends on the ratio of the standard deviation of the *sums* of the ranks to the standard deviation of a *single* set of ranks.

4. Mr. Morris also enquires about the alleged superiority of the method of paired comparison. Carried out in the manner standardized by Titchener,[2] the procedure is needlessly laborious. Most

[1] For converting ranks into normal deviates I suggest using Table XX of Fisher and Yates's *Statistical Tables* rather than Hull's table, as suggested by Mr. Morris. It may be noted that the correlation between a set of ranks and the corresponding normal deviates is quite high : with a large sample it is $\sqrt{3/\pi} = 0{\cdot}98$: (a proof is given in the *Notes* previously cited).

[2] *Experimental Psychology*. I. *Qualitative*. ii. *Student's Manual* (1901), Experiment xxi, pp. 92f.

of its advantages can be secured, and nearly all its disadvantages avoided, if the names of the items to be ranked (in the foregoing example, the names of the pupils) are typed on cards, and each card *mentally* compared with those nearest it for the quality to be judged. In the case of the eight boys mentioned above, had all the possible comparisons been explicitly made, pair by pair, and the results recorded by each teacher, we should have had a table like that shown below (Table III) ; my arrangement follows Guilford's (*Psychometric Methods*, 1936, p. 226) rather than Kendall's (*op. cit.*, p. 127). The figures show how many times the boy whose name stands at the top of the column was preferred to the boy whose name stands at the left of the row. In the diagonal I have placed (in brackets) the number of judges making the comparison, not half that number (as Guilford does), because no differences could occur between the judges about the *same* boy.

TABLE III. TEACHERS' JUDGEMENTS EXPRESSED BY PAIRED COMPARISON

Number of times boy named at top of a column was ranked above boy named at left.

Boys	T.S.	B.R.	O.F.	R.L.	E.B.	M.W.	G.A.	J.H.	Total
T.S.	(4)	3	0	0	1	0	0	0	8
B.R.	1	(4)	0	0	1	0	1	0	7
O.F.	4	4	(4)	2	3	2	1	2	22
R.L.	4	4	2	(4)	2	2	1	2	21
E.B.	3	3	1	2	(4)	3	2	2	20
M.W.	4	4	2	2	1	(4)	3	3	23
G.A.	4	3	3	3	2	1	(4)	2	22
J.H.	4	4	2	2	2	1	2	(4)	21
Total ..	28	29	14	15	16	13	14	15	144

For data of this kind a suitable method of statistical analysis has been suggested by Kendall (*op. cit.*, ch. XI).

(*a*) With the full procedure it is usually desirable to begin by testing the judgements for consistency. But, in investigations like those which Mr. Morris has in mind, inconsistencies have no interest in themselves, and merely reveal a defect in the experimental technique : either the quality has not been defined in sufficiently unambiguous terms or the judge is unintentionally departing from the definition ; and in that case it would be better to correct the inconsistent judge during the process of judging (allowing ties in case of doubt) rather than to look for inconsistencies afterwards.

(*b*) To measure the amount of agreement, Kendall proposes a coefficient which he denotes by u. Slightly simplifying his working formula, we may write :

$$u = 1 - \frac{2\,(m\Sigma p - \Sigma p^2)}{\frac{1}{2}\,m\,(m-1) \times \frac{1}{2}\,n\,(n-1)}, \qquad \ldots \text{(xiii)}$$

where p denotes the frequencies entered in the triangle below the diagonal only. Thus, with the example given above :

$$u = 1 - \frac{2\,(4 \times 30 - 54)}{6 \times 28} = \cdot 214.$$

But (as Kendall points out) u is in effect a generalization of his own coefficient for measuring the correlation between two sets of ranks, which he denotes by τ. τ, however, is always lower than the ordinary product-moment coefficient of rank correlation (ρ). Hence, to those who are more familiar with the scale of ordinary correlation coefficients, the value computed from eqn. (xiii) may be somewhat uninformative, if not actually misleading.

Since in the cases with which Mr. Morris and I are concerned the inconsistencies will be either non-existent or negligible, I venture to suggest that, for most practical purposes, it will be simplest to calculate r_1^2 and r. In experiments where either rank-methods or paired comparison may be used, the relation between the two modes of treatment can readily be seen. On comparing Tables I and III the reader will notice that, although the entries in the body of either table are different, the totals in the right-hand margin are exactly the same. A little reflection will show why this must always be so. Consider any particular boy, say J.H., the last in the list. According to Table I, Teacher i places him 3rd, and therefore ranks two boys above J.H. ; similarly Teacher ii ranks $(8 - 1) = 7$ boys above J.H. ; Teacher iii, three boys ; and Teacher iv five boys. The total of such preferences is thus $(3 - 1) + (8 - 1) + (4 - 1) + (6 - 1) = 21 - 4 = 17$. But this is the number of preferences actually recorded for J.H. in Table III (since the figure 4 in the diagonal does not represent actual preferences). Evidently then, by adding 4, we shall obtain the total entered for J.H. in Table I. Similarly for all the other boys.

It is clear, therefore, that we can use the totals obtained by paired comparison in the same way as we used the totals obtained from the ranks, to calculate the correlation ratio ; and from that we may deduce, as before, the average correlation. This method of treating the data has the advantage of keeping the constants so calculated in close relation to the results obtained from ordinary correlational techniques and in particular from a factorial analysis. Hence, except when the method of paired comparison reveals inconsistencies, there would seem to be no advantage in using the coefficient *u*.—CYRIL BURT.

THE IDENTIFICATION AND INVARIANCE OF FACTORS

By T. LEYDEN

In several papers in this *Journal* (e.g., II, pp. 134f.) various methods have been recommended for deciding how far a factor found in one research may be justifiably identified with that found in another research. The earliest suggestions appear to be those put forward by Burt, who attempted to show that the educational abilities discovered in his later surveys were the same as those reported in the earlier (cf., e.g., *J. Educ. Psychol.*, IX, pp. 68f., and refs.). He suggested (*a*) the correlation between *factor-measurements* when the samples are the same, (*b*) the ' unadjusted correlation ' between *factor-saturations* when the samples are different, (*c*) a ' symmetry criterion ' as an over-all test. More recently Fiske (*Amer. Psychol.*, III, pp. 360f.) and others have proposed (*d*) the ordinary (' adjusted ') correlation between saturations, which has proved extremely popular. Cattell has described (*e*) a ' symmetrical marker criterion,' and (*f*) a ' coefficient of pattern similarity ' derived from the formula for chi-squared (cf. this *Journal*, II, pp. 136–9, and refs.). In an investigation on which I am engaged it has been surprising to find how divergent, and even contradictory, the verdicts pronounced by these various methods frequently are, especially in the case of the first factor. Many educationists besides myself would, I know, be grateful for a fuller examination of this problem.

To check my own observations, I secured nine sets of test-data, each relating to the same group of tests (or nearly so) applied to the same group of pupils at intervals of one to three years. In three the tests were mixed batteries of intelligence-tests ; in six they were educational. Since the tests were not carried out with this particular problem in view, the results deserve only the briefest report. I hope to discuss the details more fully in a later contribution based on a more systematic enquiry planned *ad hoc*.

To compare the results obtained I have taken the correlation between factor-measurements as giving the ideal figure, and have expressed the coefficient obtained from the other procedures as a percentage of this figure.

TABLE I. COMPARISON OF CRITERIA

Criterion	Average Percentage
1. Correlation between factor-measurements	100·0
2. Correlation of factor-measurements with external criterion (teachers' assessments)	87·3
3. Unadjusted correlation between saturations	91·6
4. Adjusted correlation between saturations	14·5
5. Marker criterion	52·7
6. Coefficient of pattern similarity	61·4

The figures above relate only to the first or general factor as extracted by simple summation or the centroid method. With the corresponding ' basic ' factor, as extracted by a group factor analysis, the proportions are much the same, except in three cases where there are large supplementary factors : in such cases, not only the relative size and order of the saturations, but also the size of the factor-variances, may appreciably change in the second research. In one case, for example, it appeared that, after an interval of two years, the variance of the general factor obtained by the centroid method had apparently increased from 51 to 54 per cent., whereas the variance of the basic factor had dropped from 46 to 38 per cent. Thus the use of the centroid procedure instead of a group factor analysis may at times obscure the well-known tendency towards increased specialization of abilities with increasing years.

The main conclusion seems clear. With the general factor the popular method of correlating the two sets of correlations may yield a most misleading conclusion : (with bipolar factors, of course, the two coefficients are bound to be much the same). On the whole, the safest method seems clearly to be that based on the ' unadjusted correlation.'

BOOK REVIEWS

Factor Analysis. By Raymond B. Cattell. New York: Harper & Bros. Pp. xiv + 462. £2 8*s*.

Professor Raymond Cattell's latest book has a threefold purpose, and is divided into three corresponding sections. Part I seeks "to meet the need of the general student in science." Part II is intended to enable the volume to be used as "a text-book for statistics courses." Part III is still more advanced; and, with this further addition, the author hopes that the volume may also serve as a "handbook for the research worker."

As his bibliography shows, Professor Cattell was actively at work on factorial problems nearly a quarter of a century ago—well before other investigators in the United States began to publish contributions in this field. He received his own training at University College, London; and, as a result, unlike most American writers, he not only shows a closer appreciation of the ideas and methods of the earlier British pioneers, but is also able to combine certain features from both modes of approach. His original ideas, his long experience, and his versatile appreciation of a wide variety of devices give his pronouncements on many of the vexed questions of factorial theory an exceptional interest; and an examination of his arguments may therefore in some measure help us to decide on the relative advantages of British and American procedures and theories.

In Part I he sets out the 'Basic Concepts.' Factor analysis is described as a 'wholistic method,' which seeks to discover 'unitary tendencies,' i.e., what are popularly regarded as the 'causes' underlying 'structured groups or clusters' of correlated phenomena. In discussing the "earliest roots of modern factor-analytic ideas," he notes that "both Galton and Pearson had raised the fundamental question"—how can we ascertain and measure the common causes to which covariance is due?; and that both did much to indicate the lines along which an answer might be reached. Further, "a clear statement of multifactor analysis was given by Maxwell Garnett in 1919." Here, as Dr. Cattell incidentally observes, it is "an instructive comment on our understanding of past work" that Garnett's work should "only have been generally recognized after Thurstone had made his discovery." It might be added that Garnett's contribution is not "generally recognized" even now; and even Dr. Cattell seems to have forgotten that Dr. Garnett was Karl Pearson's assistant and explicitly attributes the essential principles to him. Of the various techniques available, we are told, "probably the principal components method—also called the principal axes method—has the greatest appeal to mathematicians." It was in point of fact the first procedure suggested for the purpose, and was developed in a lengthy classical paper published by Karl Pearson as early as 1901.[1] But "the calculations are long and difficult," and the resulting factors appear to possess "no fixed psychological meaning."

In spite of the somewhat laborious computations which any efficient method must involve, the use of some such technique (we are told) is unavoidable, because in the social and psychological sciences the investigator is nearly always forced to deal with several variables—and often with several hypothetical causes—in one and the same research. Professor Cattell therefore lays special stress at the outset on the need for the psychological student to understand the nature of 'scientific method' and the 'proper role of hypotheses.' He is, however, somewhat sceptical of the approbation bestowed by recent factorists, both in this country and abroad, on the so-called "hypothetico-deductive procedure." The logical methods by which science seeks to establish empirical relations, he says, are "essentially those set out by inductive writers like Bacon and Mill." Indeed, in his own view, factor analysis is "far more free than most methods from the necessity to elaborate rigid hypotheses: it is the ideal method of *open exploration*" (his italics), though "of course" (as he says elsewhere) "the factorist always enters with *some* hypothesis even when he seems to enter with none."

It is partly the desire for freer exploration and partly the belief that the more rigorous methods of Pearson do not yield the 'most meaningful factors' that have led to the supplementary techniques of factor-rotation. Accordingly, 'the art of rotation,' as Professor Cattell describes it, receives full treatment at an early stage. Since we should not expect any one psychological factor to affect more

[1] In Part II, curiously enough, the 'inventor' of the method is said to be Hotelling, who makes no such claim and whose version has been used even less frequently than Pearson's. During the first ten years of the century every writer who investigated the possibility of applying factorial methods to psychological problems—Spearman, Burt, and William Brown—plainly had Pearson's technique in mind, although each would also have endorsed Cattell's criticisms. The fact that Pearson's procedure has not proved practicable or appropriate in psychology without considerable modification should not allow us to overlook that it was after all Pearson who first saw the problem and produced the first multifactor technique.

than a fraction of the total personality, he himself is convinced that the most meaningful scheme will be a ' simple structure ' of group factors : such a structure, he considers, will also show most invariance.

How far a factor is invariant must depend, as he rightly points out, on many more conditions than are commonly assumed. Hitherto invariance has usually been interpreted to mean stability of factors, when *tests* are varied. But the psychologist's initial measurements, like the physicist's, imply variation in three distinct ' dimensions.' As Stern pointed out long ago, they consist of assessments for certain *attributes* obtained from certain *individuals* at certain *times*. Hence it would seem possible, *a priori*, to correlate and factorize either attributes, or times, or individuals. And in some of the earliest British investigations we find profiles derived to describe individual children, and factorizations of test-results obtained on successive days of the week or at different stages of the child's growth. Spearman and Thurstone, on the other hand, seldom mention either the individuals as such or the changes they exhibit with the passage of time ; for the most part they and their followers treat factor analysis as concerned primarily with attributes, and chiefly with cognitive attributes. Professor Cattell agrees with those who adopt a wider standpoint. He ingeniously elaborates the box form of Stern's three-dimensional chart, and uses it to provide a somewhat elaborate symbolism to distinguish the various conceivable approaches. Thus " R1B-technique is the typical procedure ; Q2B-technique is Burt's advocated use (of correlating persons) ; Q1S is Stephenson's ; and P1B the analysis of the single individual." As Cattell remarks, what might seem at first a complex and confusing situation was greatly simplified by Burt's demonstration of the ' reciprocity principle.' Since in theory the matrix factorized by one method is simply the ' transpose ' of the matrix factorized by the other, the factors obtained by one technique must in theory be the same (except for those eliminated by preliminary averaging) as the factors obtained by the other : e.g., to take Cattell's illustration, " if correlation of abilities yields a factor of mechanical aptitude, then correlation of persons will yield a factor or type consisting of mechanically apt persons," and so on. The theorem is really a special instance of Poncelet's ' duality principle,' which has proved so fruitful in projective geometry. This conclusion, although at first vigorously attacked, is, so Cattell considers, fully confirmed by the recent work of Madow. It would seem to follow, as he says, that the views still held by Stephenson about the independence of Q- and R-technique cannot be accepted. In practice of course there will be many minor differences arising from the different modes of approach ; but Cattell's final conclusion is that " R-technique provides the binding frame of reference."

Part II deals with ' Specific Aims and Working Methods.' Professor Cattell begins by classifying the various working procedures available. " Of the possible classifications according to the algebraic process, the division into the methods of *simple summation* and *weighted summation* is one which is fundamental for mathematicians and also important for the computer." But there is a further distinction, no less essential to an understanding of the conclusions reached, between what he prefers to call ' bipolar solutions ' (i.e., summational and ' centroid ' factors) and ' bifactor (i.e., general and group factor) solutions.' [1] The former, he agrees, might be regarded as " difference factors with respect to the latter " ; but he doubts whether this is always true. On the whole he considers that " the most likely configuration is one of overlapping group factors with any combination of signs that may be required," in short, a simple multifactor structure which may have negative as well as positive loadings. " The real objection," he says, " to Burt's argument for ' bifactor ' methods is that the factors obtained are not the same in meaning as those obtained from simple structure centroid analysis " : even if the ' primary abilities ' reached by Thurstone—the verbal, spatial, numerical, and other group factors—appear to be similar to those previously found by Burt, there is nevertheless " *one* factor in the bifactor series, namely, the first general factor, which is *never* matched by anything in the multifactor series " (his italics). But surely this argument will cut both ways. A critic might equally well maintain that the oblique factors obtained in a ' simple structure ' are " not the same in meaning " as those obtained by an analysis into general and group factors, and that they omit

[1] This use of terms is likely to be a little confusing to the beginner. The terms ' bipolar,' ' bifactor, and ' centroid ' are so closely associated with the methods of the writers who introduced them that it would seem safer to allow each a kind of copyright for his own specific terms. On p. 133 Cattell appears to suggest that the scheme of general-plus-group factors was a later modification of " the process called bifactor analysis by Holzinger," and was " conceived as a direct extension of Spearman's two factor solution." Spearman, however, explicitly states that Burt's method was " chronologically earlier than the bifactor method " (*Human Ability*, p. 28) ; and surely the phrase ' bifactor method ' is needed to describe the *particular working method* devised by Holzinger and his co-workers. The resulting factor-matrices may appear the same in their *general* structure, but in their details the factors and their specifications may differ appreciably. For example, one of Cattell's objections to the ' bifactor solution ' is that the results are ' not invariant ' : this may be true of the results reached by Holzinger's procedure for the reasons he gives ; but it does *not* appear to be true of the scheme of factors reached by the so-called ' group factor method.' Indeed, owing to the objectivity of the latter procedure, the results are much more likely to be invariant than those of simple structure, owing to the wide choice of " alternative sets " to which Cattell himself alludes.

the factor which, as a rule, accounts for quite as much of the variance as all the group factors put together. It seems to have been a belated recognition of this point that led Thurstone to suggest re-factorizing the correlated factors, and so securing a ' second order ' matrix, containing a general factor as well as group factors.

Cattell's own view of simple structure, it will be noticed, departs in one important respect from that of Thurstone ; he does not insist that it should contain positive saturations only. " The myth that a positive manifold must be maintained," he says, " we have already dismissed, since almost any factor is as capable of interfering with a performance as of aiding it." He mentions further difficulties attending the search for simple structure. " The discovery that one and the same correlation matrix may yield more than one simple structure solution, and therefore *alternative sets of factors*, presents a disturbing problem." To secure a decisive result " a common device among the unsophisticated is to rotate for psychological meaning, i.e., to move a factor to a position where the loadings agree with some preconception of the experimenter." But, he continues, " almost any preconception can be confirmed in this way, and this approach merely perpetuates erroneous speculations."

To escape from this situation various devices are considered. One obvious solution, as he points out, is " not to rotate at all ! The most systematic defence of this procedure has been made by Burt, followed by Eysenck." But is this a fair representation of these writers' views ? Burt, for example, as readers of this *Journal* will be aware, has described both arithmetical and graphical procedures for transforming a set of summational (or ' bipolar ') factors to a corresponding set of general and group factors ; and, in order to avoid the subjective influences, like those of which Professor Cattell complains, the rotations are based on the dichotomous classification indicated by the bipolar factors. Eysenck, in the earliest of the papers cited by Cattell,[1] follows (as he says) an alternative method " used by Burt and his co-workers nearly twenty years ago " ; this is to " make graphs of the [observed] coefficients and judge by the resemblances between the contours " : tests that yield parallel contours are then assigned to the same group factor. It is true that there is here no explicit rotation : but the purpose of the calculations was to show that, with certain types of correlational table, what are virtually ' rotated factors ' can be reached without " the prolonged and admittedly precarious graphical procedure." In his later work rotation has often been employed.

Elsewhere Cattell himself describes what would seem to be an extension of much the same principle—the use of " parallel proportional profiles." This he would regard as " the ideal method " ; but, he adds, " it is not at present in a practicable state," and moreover " for simple structure it has several drawbacks." As a conceivable alternative, he discusses ' criterion analysis,' or, as he prefers to call it,[2] ' criterion rotation ' ; and adduces " logical objections to regarding it as an independent or sufficient method." In particular, he contends, since it seeks to make the so-called factor " correspond exactly " to the criterion, it is not really a method of factor analysis at all. After a long and instructive discussion of various ' specialized techniques for rotation,' he finally concludes that " bringing the rotation process to a satisfactory conclusion involves more than a mechanical pursuit of simple structure : it requires an appreciation of the meaning of the constellations."

Part III is entitled " General Principles and Problems " ; and " is concerned with wider theoretical vistas and with special refinements of practical aids." Here he echoes a plea so often made in the pages of this *Journal*, namely, that, instead of experimentalists and factorists each remaining indifferent or hostile to the complementary approach, efforts should be made to combine the two. What is still more welcome, he puts forward several ingenious suggestions to show how such attempts might proceed.

All Professor Cattell's pages have a racy liveliness that marks the quick and original writer. But some of them at times betray signs of haste. There are slips and obscurities, not only in the phrasing, but also in the figures : e.g., the reference on p. 249 should be ' 100a,' not ' 153,' and the reference on p. 315 should be ' 1a ' ; the ' correlation matrix ' referred to on p. 202 will be found on p. 41, not, as stated, on ' page 46 ' ; on p. 165 question 4 sets the student an exercise in calculating centroid factors, but the ' solution to the question ' on p. 166 gives figures that would not be obtained in a centroid analysis, and the factor-variance for the second factor proves to be more than twice as large as that for the first ; Table 6 contains a puzzling misprint.

[1] ' Critical Notice of Thurstone's *Primary Mental Abilities*,' *Brit. J. Educ. Psychol.*, IX, 1939, p. 272. A closer study of this paper, which attempts a re-analysis of Thurstone's own correlation table, scarcely bears out Cattell's own conclusions about the superiority of Thurstone's methods. Of the common factor-variance, nearly 60 per cent. can be accounted for by a single general factor, while only 40 per cent. is covered by the seven group factors corresponding to those in Thurstone's " simple structure of primary abilities." Possibly, however, Cattell himself would here agree, since he states that " for some purposes there is little to choose between the two formulations—the rotated centroid yielding primaries and the bifactor yielding a large general factor plus what is left of the primaries."

[2] The reason is that " the term *criterion analysis* has been already used in a more appropriate sense by other factor anaylsts."

Inaccuracies and obscurities of this kind, however trivial, are unfortunate in those chapters which provide the student with his first introduction to the working procedure. In Table 1 Professor Cattell sets out the correlation table, which is to serve throughout the book as his ' main example ' ; and, since his treatment is fundamental to his general exposition, it is worth while examining it in some detail. In ch. III he explains how the student is to extract the first factor by what he calls the ' centroid method,' which, he says, was introduced by Thurstone. Yet both the formula cited and the procedure actually followed are, in point of fact, those which had been used by British investigators long before Thurstone introduced the ' centroid method ' ; and, indeed, Thurstone's modifications do *not* seem to have been adopted. In particular, the method of assessing communalities [1] and the method of reflecting signs depart a good deal from those which are the distinctive features of Thurstone's working procedure.

In the next chapter the student is told how to extract the remaining factors in order of significance until no significant figures are left. But here something seems to have gone wrong either with the account of the procedure or with the way it has been carried out. First, as a little calculation quickly shows, the variance of the second factor is only 0·987, while that of the third is 1·148 ; secondly, in this last factor the saturations for variables in the negative half of the preceding factor yield positive and negative totals that are far from equal (+·81 and −·65) ; and finally if we try to reconstruct the observed matrix, the fit is by no means satisfactory. These anomalies appear due to various causes. (1) To begin with Professor Cattell has not noticed that in two places the communalities he has used *destroy the Gramian character* of the correlation matrix, and hence are bound to lead to inaccurate figures. (2) In dealing with what he calls " the exasperating business of reflecting," his own selection of variables for sign-reversal is by no means the best ; and it is this that accounts for the low variance of the second factor.[2] (3) He tells the student that, since the observed coefficients are given to no more than two decimal places, " all the figures in the computation " should be rounded off to two places : but, if only two figures are retained in the computation, the final saturations are likely to show rather large errors in the second place. (4) The method of allocating the signs to the saturations for factors I and II in Table 8 gives *negative* values to those tests which (as even a casual inspection of the original correlation table shows) must have *positive* values in the group factor matrix. It is, therefore, hardly surprising that Professor Cattell is induced to conclude that his centroid factors are devoid of meaning. But of course he might retort that at this stage he is not concerned with meanings, but only with obtaining something to rotate.

In ch. V the student is told ' how to rotate for scientific meaning.' Cattell's Table 9 gives the " orthogonally rotated factor-matrix " reproduced below, and appends new communalities " obtained from the actual loadings by squaring " : (incidentally, the squared loadings for test 4 do not add up to the communality stated, namely, 0·98). The reader is then assured that " further rotation fails to improve on the above," and that " the best simple structure has been obtained." Yet on a later page (Table 27) some of the values are improved, without any specific indication of the change. But even so, when the captious reader tries to reconstruct the initial correlations he quickly discovers that there is still room for a good deal of further improvement.

To demonstrate these points, let us examine the results obtained from a straightforward application of the method of simple summation. Let us adopt the method of ' simple summation ' in its original form [3] and then apply an arithmetical rather than a graphical rotation. One iteration only

[1] I have found more than one of my colleagues baffled by the attempt to discover how the proposed communalities were reached. In the note to Table 1 (p. 41) the student is told that " the method by which they are ' guessed ' will be explained later " ; but nothing more is said about the matter in this chapter. When he reaches ch. X, he is told that " a rough method—No. 1 below—has already been given in chapter III " (' No. 1 below ' consists in taking the highest coefficient in the column, but rounding it off to one significant figure only). But this method has *not* been given in ch III ; and it is plainly *not* the method that has been used in that chapter : (in cols. 4 and 5 " the highest coefficient in the column " is ·66, but the figure inserted is ·90 ; in cols. 7 and 8 " the highest coefficient " is ·72, but the figure inserted is ·80 in col. 7 and ·60 in col. 8). A later page suggests rearranging the variables " in order of their mean r's," and considering the general trend instead of just blindly inserting the highest. " These considerations " (it is said) " lead to the practice suggested by Burt," and usually mean " estimating communalities with high mean r's to be higher than their highest r's, and those of low mean r's to be a little lower." This method is called ' No. 2,' and would seem to be the method actually adopted in ch. III ; but it is certainly not " that introduced by Thurstone."

[2] Even without rearranging the figures, an inspection of the table of ' first residuals ' seems sufficient to indicate the kind of sign pattern required : by far the largest figures are those in the column for test 5, and here the positive and negative signs plainly contrast tests 2, 4, and 5 with tests 1, 3, 6, 7, and 8. Having decided (presumably from the positive block at the corner) to group 6, 7, and 8 together, it is extremely difficult to see why Dr. Cattell then groups 5 with these.

[3] Cf. Burt, C., *Factors of the Mind*, pp. 461f., or earlier papers. The working was carried to four decimal places.

was needed to reach satisfactory communalities ; and with these it appeared that all the correlations except the last two could be completely accounted for by three factors only : the last two required a small supplementary factor. The saturations are shown in Table I A. Using a strict arithmetical rotation (based on the procedure described in this *Journal*, III, p. 75) we then obtain the rotated factor-matrix shown in Table I B. These factor-saturations enable us to reproduce the correlations *exactly*. I do not know what may be the source from which Professor Cattell derived his correlation table ; but it is tempting to hazard a guess that these were probably the figures used to construct it. The results of Professor Cattell's own rotation are inserted in Table I C. It will be seen that as many as seven of the saturations printed deviate by between ·10 and ·15 from what may be regarded as the true values : they are deviations which, with a sample of moderate size, might well exceed twice the standard error.

TABLE I. FACTOR-SATURATIONS

Method	A. Simple Summation				B. Group Factors				C. Rotated Centroid Factors (Cattell)		
Test	I	II	III	IV	I′	II′	III′	IV′	F_1'	F_2'	F_3'
V_1	·344	−·123	·486	·000	·0	−·1	·6	·0	−·05	−·61	·10
V_2	·396	·341	·217	·000	·0	·4	·4	·0	·46	−·37	−·04
V_3	·110	−·384	·401	·000	·0	−·4	·4	·0	−·38	−·47	·13
V_4	·932	·337	·167	·000	·4	·6	·7	·0	·75	−·57	·38
V_5	·528	·688	−·384	·000	·3	·9	·0	·0	·85	·08	·17
V_6	·033	−·133	−·036	·000	·1	−·1	·0	·0	−·11	·01	·14
V_7	·682	−·385	−·455	·100	·9	·0	·1	·1	·12	·02	·89
V_8	·613	−·342	−·396	−·100	·8	·0	·1	−·1	·10	·04	·80
Sum of Squares	2·276	1·148	·986	·020	1·71	1·51	1·19	·02	1·68	1·06	1·65

In case the reader cares to check the calculations, Table II shows the rotation matrix obtained arithmetically. Since Cattell's bipolar-matrix is based on somewhat unsatisfactory estimates of the common factor variances for the several tests, it is not possible to obtain a very good approximation to the ' true ' values (particularly if an orthogonal transformation is used) : I give a transformation based on a least squares fit, and append Cattell's rotation matrix for comparison.

TABLE II. TRANSFORMATION MATRICES

Arithmetical					Graphical						
A. Based on Loadings in Table I					B. Based on Cattell's Loadings				C. Cattell's Rotation Matrix		
	I′	II′	III′	IV′		I′	II′	III′		F_3'''	F_1'''	F_2'''
I	·6873	·3574	·6323	·0000	F_1	·7093	·3858	·6202	F_1	·71	·53	−·47
II	−·4213	·9054	−·0538	·0000	F_2	·6165	·1236	−·7643	F_2	·50	·10	·86
III	−·5917	−·2293	·7729	·0000	F_3	·3324	−·9141	·1515	F_3	·50	−·84	−·19
IV	·0000	·0000	·0000	1·0000		. . . . .				. . . . .		

It will be agreed, I think, that the bipolar-matrix is here by no means devoid of meaning : on the contrary, it clearly indicates the location of the various group factors : the first bipolar (factor II) suggests a group factor with positive saturations for tests 2, 4, and 5 (such as are found in factor II′) ; the second bipolar (factor III) indicates a group factor with positive saturations in tests 1, 2, 3, and 4 (such as are found in factor III′) ; the supplementary factor (factor IV) remains unaltered ; and finally factor I suggests a factor with positive saturations in tests 4, 5, 7, and 8, and implies an appreciable overlapping in tests 1, 2, and 3.

Some psychologists will be inclined to ask whether these peculiarities in working methods may not be responsible for some of the rather unexpected factorizations that Dr. Cattell has at times obtained for his own empirical data (cf., for example, this *Journal*, II, pp. 114-30, and Dr. Banks' criticisms and re-analysis in II, pp. 204-18) ; and possibly a good many will feel that the artificial table thus chosen as the main illustration is scarcely typical of those which the majority of students will encounter in their work. Moreover, if the student possesses some knowledge of mathematics already, he would himself probably wonder why the elaborate calculations, described stage by stage in ch. III, IV, V, X, and XII, are needed for this table. If, in addition, he has already gained experience in group factor procedures, he would omit both the preliminary centroid analysis and the prolonged rotations, and obtain the final figures (those of Table II B) at the very first step.

At the same time it would be unwise to press these criticisms too strongly. Hitherto most text-books on factor analysis have been chiefly concerned with the factorization of human abilities ; Dr. Cattell's book is welcome because he has adopted a much broader standpoint and has throughout kept well in mind the factorization of human personality generally. It is therefore conceivable that, in a fuller and more technical discussion, he might be able to adduce instructive arguments in favour of what at first sight seem rather inappropriate procedures.

In any case, whether the reader accepts these methods and whether he agrees with the doctrines expounded or not, it is impossible to deny that the book itself provides one of the most stimulating introductions to the subject, and is full of novel, original, and often highly illuminating suggestions.

E. N. Harper

Tests Collectifs du Centre de Recherches. By M. Reuchlin and E. Valin. Paris : Institut National d'Étude du Travail et d'Orientation Professionnelle. 1953. Pp. 152. 400 *fr.*

This monograph presents a study of the battery of tests compiled by the Research Centre of the *Institut National d'Orientation Professionnelle*. The first part summarizes the main practical conclusions ; the second and longer part deals with the more theoretical questions of methodology. Here the authors have " devoted an important place to certain procedures which seem of special interest for such work, but have nevertheless been but little used in France," namely, the procedures usually called ' factor analysis,' " to which," they say, " our analysis appears to make some contribution."

" It was," they explain, " between 1910 and 1920 that Burt found that correlations between cognitive tests could most frequently be accounted for by two types of factor—a general factor common to all and a series of group factors each common only to a limited number : this hypothesis was opposed by Spearman, who at first refused to admit the existence of group factors ; but it is equally opposed to the analysis into ' simple structure ' developed later on by Thurstone." However, M. Reuchlin and M. Valin take the view that, in the latter case at any rate, the apparent conflict is by no means irreconcilable ; and they consider that their own figures clearly demonstrate that " the results obtained by Thurstone's method (provided it is pushed to its final conclusion) lead to those obtained more simply by the method of analysis already described by Burt." It would be instructive to hear how far Thurstone and his followers accepted this conclusion.

For the purposes of their main enquiry a battery of 15 tests was administered to nearly 2,000 boys and girls aged approximately 14 years. The children were divided into four subgroups—town boys, country boys, town girls, and country girls. A table of correlations was calculated for each group and then factorized by various procedures. Twenty-six pages of tables give their results in some detail.

The writers begin with an ordinary ' bipolar analysis ' based on the principle of ' simple summation.' This leads them first of all to discuss the similarities and differences between Burt's ' simple summation method ' and Thurstone's ' centroid method.' The most important divergence, in their view, arises from the way the reduced self-correlations (or ' communalities ') are assessed : Burt's method is to estimate the self-correlations by inspecting the general trend of the observed correlations (not relying *solely* on the largest coefficients), and then to correct these provisional estimates by the results obtained from an actual analysis, repeating the whole process until stable figures are reached. Thurstone, on the other hand, simply inserts the largest figure from each column, both when extracting the first factor and when dealing with the later factors. The French investigators consider that the procedure adopted by British workers " is much more satisfying in principle, but obviously more exacting to apply." With the present data they find about four iterations usually sufficient ; and their final calculations, they believe, warrant the extraction of four factors.

The bipolar factor-matrices reached by this procedure they regard as somewhat difficult to interpret, because " the negative saturations cannot easily be justified psychologically : in the cognitive field it is not easy to imagine two groups of tasks of such a nature that success in one group necessarily leads to relative failure in the other." This argument, however, assumes that the factors obtained are to be identified, just as they stand, with cognitive abilities. But according to Burt—and the view has earned wide acceptance—factors are primarily principles of classification, and bipolar factors are primarily principles of dichotomous classification. On this basis the results can

at once be interpreted. In three of their groups, the first bipolar factor evidently contrasts tests requiring abstract or intellectual processes with those requiring more concrete or practical processes ; in the fourth group the same division is effected by the second bipolar factor. In each case the other bipolar factor contrasts contents—namely, verbal with numerical and spatial with logical.

To secure an interpretation in terms of positive abilities, the authors next proceed to ' rotate ' the initial factor-saturations, first in accordance with the methods described by Thurstone and secondly in accordance with those originally proposed by Burt.

In Thurstone's procedure, the first stage is to obtain a ' bipolar-matrix ' of ' centroid factors.' The second stage consists in applying a succession of graphical rotations until some kind of ' simple structure ' is achieved, which will consist of group factors only with no general factor. The investigators offer very convincing evidence for regarding the results so reached as inadequate if not actually misleading. It is true that, by permitting ' oblique ' or correlated factors, Thurstone usually succeeds in avoiding anything like a general factor ; but this, it is argued, leaves readers quite in the dark about the extent to which the various tests chosen for the battery really tend to measure the same thing : in short " they are kept in the same state of ignorance, and therefore exposed to the same kinds of error, as in the days when factor analysis had not been developed, and test-results were commonly interpreted by mere inspection." Thurstone, we are told, in a desire to simplify the naïve scheme of general and group factors (i.e., of ' general ability ' and ' special aptitudes '), attempts to throw the entire onus of explanation on group factors or special aptitudes only, just as Spearman at one time hoped to explain everything by appealing to a general factor only, i.e., the factor of ' general intelligence.' But, it is argued, since the general or basic factor contributes quite as much to the total variance as the group factors do, the new type of ' parsimony ' seems just as indefensible as the old. Why, simply on the grounds of personal preconceptions, should we ignore the useful general factor any more than we ignore the narrower group factors ? The conclusion therefore which the authors draw is that " whenever oblique factors are used for the first order analysis, a second order analysis is absolutely indispensable."

At this third stage, as they very justly assert, it is not sufficient to calculate just *one* general factor of second order ; the saturations for *all* the second order factors in all the tests should be computed. Unfortunately " no American author using second order factors has, to our knowledge, pushed the analysis as far as this, although the necessary technique permitting this has been described by Thurstone." At this stage it will once again be necessary to begin by estimating the diagonal values or ' self-correlations ' : for this purpose, they say, " nous avons estimé d'abord les communautés par ' smoothing,' suivant la méthode décrite à propos de l'analyse par ' sommation simple ' de Burt, et ensuite procédé à des réestimations successives." The final outcome of the Thurstone analysis consists in four factor matrices—one for each batch of children : they include, besides the first or ' basic ' factor, three supplementary in the case of the town boys and town girls, four with the country girls, and five with the country boys. In the last case the additional factor really forms a second general factor ; its appearance is attributed to the fact that (as not infrequently happens with Thurstone's procedure) the correlations between the oblique factors would lead to self-correlations over 1·000, if a single general factor alone were used. In the main the supplementary group factors thus obtained are very similar in nature to those which appear in the ' simple structures ' : but they have the advantage of being orthogonal instead of oblique or correlated. The inter-correlations both of the tests and of the first order factors are now expressed—and better expressed—by a general factor. It is consequently inferred that " analyses into first order oblique factors, which dispense with a general factor, cannot of themselves provide a satisfactory foundation either for theory or for practice."

The writers turn next to what they call the ' méthode de Burt.' In the main they follow the procedure outlined in the article on ' Group Factor Analysis ' in this *Journal* (III, pp. 40–75). They agree that this procedure is far less subjective than that of Thurstone, but contend that the subjective element is still not wholly eliminated. British factorists commonly assume that, if a summational analysis is carried out first and the resulting sign-pattern adopted as a guide (to be superseded of course by the results of successive approximation), then the final subdivision of the tests or traits into groups and subgroups emerges automatically. Drs. Reuchlin and Valin feel that, owing to the lack of any test of statistical significance, there may still be room for arbitrary decisions, when the factor-saturations are extremely small. Moreover, in calculating the preliminary matrix of non-overlapping group factors, they are a little dismayed to find that somewhat different cross-classifications appear with the different groups of children ; for instance, two of the tests involving number are classed sometimes with the tests requiring numerical ability and sometimes with the tests of reasoning. But, since each test involves both processes, and since boys and girls (or urban and rural pupils) may conceivably make use of either process in varying degrees, it is not surprising to find such discrepancies ; indeed, so far from constituting defects of the method, they probably indicate genuine differences in the way different pupils tend to attack such problems.

Where there is any doubt, however, the *full* factorial procedure should, if possible, be followed, and then every decision will be determined by the data. If, for example, during the initial stages

a low negative saturation leads to an incorrect grouping, this will be mechanically corrected when the non-overlapping factors are replaced by overlapping factors. No doubt in practice these elaborate precautions are seldom employed. Unless the numbers tested are extremely large, the minor changes that would result could, as a rule, claim only moderate probability, and consequently it seems legitimate to shorten the procedure, as indeed the French investigators have done. Where a good many iterations are required, they point out that the procedure can be greatly facilitated, "en procédant, comme Burt le fait souvant, par *condensation*," or by graphical devices. But probably most psychologists would hold that, whenever the doubtful points are of importance, the wiser course would be, not to spend time on more refined calculations, but to plan a new and more suitable experiment to solve the issues so raised.

At the later stages, where rotations are required to show the extent of overlapping, the French writers prefer (with most British investigators) to rely mainly on an arithmetical method of rotation. For the final figures, however, instead of continuing the iterative procedure when only slight modifications are likely to result, they have " effected a few slight orthogonal rotations with the graphical procedure." In doing so, they are inclined to wonder whether their procedure may not be " une utilisation maladroite," departing a little too far from the rigorous ideal. But, as Burt and others have observed, provided the rotations are carried out ' blindly ' and in accordance with perfectly objective principles, graphical procedures may be used (by those who prefer them) even when the computer wishes to obtain a matrix of general and overlapping group factors which are perfectly orthogonal. Though seldom so accurate, they are occasionally speedier than arithmetical iterations.[1]

The results at which the investigators finally arrive are highly instructive. All the factor-matrices clearly reveal the tendency towards " the recurrent pattern of *g*, *v*, *n*, and *k* " which so often appears in British work. In the tables obtained from the country boys, there is first a general or basic factor accounting for nearly 70 per cent. of the common factor-variance (rather over 40 per cent. of the total variance) and four group factors—verbal, numerical, spatial, and (the smallest of all) reasoning. Moreover, in these instances there is practically no overlapping. In two groups, however, the reasoning factor shows a tendency to fuse with the numerical and the spatial factors. In the results obtained from the country boys and from the country girls, there is a large group factor covering the ' intellectual ' tests (as distinct from the ' practical '), and this then tends to subdivide into separate verbal and numerical factors which do not overlap.

But the finding on which the writers lay greatest stress is the remarkable similarity in every group between the factor-matrices obtained with Burt's relatively simple methods and those eventually reached with Thurstone's second-order analysis when carried through all four stages to its logical conclusion. The two methods therefore do not really lead to inconsistent results as is so often supposed : a ' reconciliation ' is possible. Further, " il nous parait important et équitable de souligner que cette reconciliation n'a pas exigé que Burt modifie sa conception première : ce sont ses ' contradicteurs ' qui se rangent, ou semblent devoir se ranger, à son avis."

P. Durant

[1] In such cases the graphical method adopted is not quite the same as that used by the French authors. Instead of the first arithmetical rotation, based on non-overlapping group factors, the computer usually begins with a preliminary arithmetical rotation by a standard *transclassificatory operator*. (This is a special type of ' equilibrated orthogonal matrix ' : for its nature and use, see this *Journal*, I (1947), p. 8.)

It may be noted that, in what would seem to be the earliest published example of factor-rotation, Garnett adopted a simple graphical procedure (*Proc. Roy. Soc.*, A, XCVI, pp. 91f.). Burt also used it in factorizing the first large American investigation on the ' Correlation of Mental Abilities ' (Simpson, B. R., *Columbia Contributions to Education*, 1912). Indeed, it was Spearman's subsequent criticisms of the alleged group factors found in this table (*Abilities of Man*, 1925, pp. 145–6, 151) that demonstrated the need for a more rigorous procedure or establishing such factors.

INDEX TO VOLUME VI

INDEX OF AUTHORS

Authors	*Titles*	*Page*
Burt, C.	Scale Analysis and Factor Analysis	5–24
Burt, C.	The Relative Merits of Ranks and Paired Comparisons	112–119
Dunsdon, M. I., and Roberts, J. A. Fraser	The Relation of the Terman-Merrill Vocabulary Test to Mental Age in a Sample of English Children	61–70
Ehrenberg, A. S. C.	On Making Statistical Assumptions	41–42
Eysenck, H. J.	The Scientific Study of Personality	44–46
Gourlay, N.	Covariance Analysis and its Applications in Psychological Research	25–34
Guttman, L.	A Note on Sir Cyril Burt's 'Factorial Analysis of Qualitative Data'	1–4
Leyden, T.	The Identification and Invariance of Factors	119
Morris, R. E.	The Relative Merits of Ranks and Paired Comparisons	112
Richardson, L. F.	The Submissiveness of Nations	77–90
Richardson, L. F.	On the Obligation to Refute Rival Hypotheses	107
Roberts, J. A. Fraser, and Dunsdon, M. I.	The Relation of the Terman-Merrill Vocabulary Test to Mental Age in a Sample of English Children	61–70
Scarr, Elizabeth H.	Changes in Terman-Merrill I.Q.s with Dull Children	71–76
Sen, Amya	A Preliminary Study of the Thematic Apperception Test	91–100
Slater, P.	The Factor Analysis of Negative Correlations	101–106
Whitfield, J. W.	The Distribution of Total Rank Value for One Particular Object in *m* Rankings of *n* Objects	35–40
Wrigley, J.	The Publication of Tables	52

INDEX OF SUBJECTS

Titles	*Authors*	*Page*
Covariance Analysis and its Applications in Psychological Research	Gourlay, N.	25–34
Factors, The Identification and Invariance of	Leyden, T.	119
Hypotheses, On the Obligation to Refute Rival	Richardson, L. F.	107
Nations, The Submissiveness of	Richardson, L. F.	77–90
Negative Correlations, The Factor Analysis of	Slater, P.	101–106
Personality, The Scientific Study of	Eysenck, H. J.	44–46
Qualitative Data, 'A Note on Sir Cyril Burt's Factorial Analysis of'	Guttman, L.	1–4
Ranks and Paired Comparisons, The Relative Merits of	Morris, R. E.	112
Ranks and Paired Comparisons, The Relative Merits of	Burt, C.	112–119
Rank Value, The Distribution of Total, for One Particular Object in *m* Rankings of *n* Objects	Whitfield, J. W.	35–40
Scale Analysis and Factor Analysis	Burt, C.	5–24
Statistical Assumptions, On Making	Ehrenberg, A. S. C.	41–42
Tables, The Publication of	Wrigley, J.	52
Terman-Merrill I.Q.s, Changes in, with Dull Children	Scarr, Elizabeth H.	71–76
Terman-Merrill Vocabulary Test, Relation of, to Mental Age in a Sample of English Children	Dunsdon, M. I., and Roberts, J. A. Fraser	61–70
Thematic Apperception Test, A Preliminary Study of the	Sen, Amya	91–100

BOOKS REVIEWED

Authors	*Titles*	*Page*
Brower, D., and Abt, L. E.	Progress in Clinical Psychology	55–56
Burt, C.	Contributions of Psychology to Social Problems	57
Burt, C.	The Causes and Treatment of Backwardness	58
Cattell, R. B.	Factor Analysis	120–125
Ferguson, T.	The Young Delinquent in his Social Setting	56–57
Fisher, R. A.	Contributions to Mathematical Statistics	53–54
Reuchlin, M., and Valin, E.	Tests Collectifs du Centre de Recherches	125–127

PR | G2 —

Vol. VI Part I May 1953

THE BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

EDITED BY
CYRIL BURT

WITH THE ASSISTANCE OF
CHARLOTTE BANKS AND ARTHUR SUMMERFIELD

AND THE FOLLOWING EDITORIAL BOARD

A. C. AITKEN
M. S. BARTLETT
W. G. EMMETT
M. G. KENDALL
D. N. LAWLEY
E. S. PEARSON
L. S. PENROSE
J. FRASER ROBERTS
A. RODGER
W. STEPHENSON
P. E. VERNON
J. W. WHITFIELD

Managing Editor : H. G. MAULE

UNIVERSITY OF LONDON PRESS LTD.

WARWICK SQUARE, LONDON, E.C.4.

Price Fifteen Shillings

CONTENTS

		Page
GUTTMAN, L.	A Note on Sir Cyril Burt's 'Factorial Analysis of Qualitative Data'	1
BURT, C.	Scale Analysis and Factor Analysis: Comments on Dr. Guttman's Paper	5
GOURLAY, N.	Covariance Analysis and its Applications in Psychological Research	25
WHITFIELD, J. W.	The Distribution of Total Rank Value for One Particular Object in *m* Rankings of *n* Objects	35
NOTES AND CORRESPONDENCE	. . . . .	41
BOOK REVIEWS	. . . . .	53

This *Journal* is issued by the British Psychological Society. The subscription price is 30*s*. a volume. Members of the Society should send their subscriptions to THE SECRETARY, British Psychological Society, Tavistock House, London, W.C.1 ; non-members to THE MANAGER, University of London Press Ltd., Warwick Square, London, E.C.4, or to any bookseller.

In accordance with proposed modifications in the titles of the journals which it publishes, the Society has decided that, in future, this *Journal* shall be known as *The British Journal of Statistical Psychology*. Owing to the increased cost of paper and printing, the Society regrets that, until further notice, it will be possible to issue only two parts each year, namely, in May and November. The size of each volume will therefore be about 130 pages. Since the number of subscribers is increasing, it is hoped that in the near future it may be possible to increase the number to three parts annually.

Papers for publication should be sent to Sir CYRIL BURT, 9 Elsworthy Road, London, N.W.3. Contributors will receive 25 copies of their papers free. Extra copies may be purchased if the order is given when the proofs are returned. Contributors should indicate whether they wish their reprints to be bound in covers. The publishers will state the current prices when sending out proofs.

Authors are asked to send their papers in a form which is ready for submission to the printer : that is to say, the arrangement of the manuscript (including a brief list of contents at the head and a summary of results and conclusions at the end) should follow the general arrangement observed in the publications of the British Psychological Society.

Books for review should be sent to one of the Editors ; advertisements for insertion to the Manager, University of London Press, Ltd.

The British Psychological Society also issues quarterly *The British Journal of Psychology, General Section*, and *The British Journal of Medical Psychology*. Subscriptions, 40s. per annum for either journal, should be sent to the Cambridge University Press, Bentley House, Euston Road, London, N.W.1. The Society also publishes, jointly with the Training College Association, *The British Journal of Educational Psychology*. Subscriptions, £1 per annum, for this Journal should be sent to Methuen & Co., Ltd., 36 Essex Street, London, W.C.2.

Papers for publication in the *General Section* should be sent to Professor D. W. HARDING, the Psychological Laboratory, Bedford College, London, N.W.1 ; those for publication in *The British Journal of Medical Psychology* should be sent to Dr. J. D. SUTHERLAND, 2 Beaumont Street, London, W.1 ; those for publication in *The British Journal of Educational Psychology* to Emeritus Professor C. W. VALENTINE, The White House, Wythall, Nr. Birmingham.

HUMAN RELATIONS

A Journal of Studies towards the Integration of the Social Sciences

Contents of VOL. VI

No. 1

ELLIOTT JAQUES. On the Dynamics of Social Structure.

JAY M. JACKSON. The Effect of Changing the Leadership of Small Work Groups.

RONALD TAFT. The Shared Frame of Reference Concept Applied to the Assimilation of Immigrants.

BEN WILLERMAN and LEONARD SWANSON. Group Prestige in Voluntary Organizations : A Study of College Sororities.

J. M. M. HILL. A Note on Labour Turnover in an Iron and Steel Works.

Book Reviews.

No. 2

CHARLES BIRD, ELIO D. MONACHESI, and HARVEY BURDICK. Studies of Group Tensions. II : Methods for Validating a Questionnaire concerning White-Negro Relationships by means of Voluntary Statements.

P. G. HERBST. Analysis and Measurement of a Situation : The Child in the Family.

P. H. COOK and A. J. WYNDHAM. Patterns of Eating Behaviour : A Study of Industrial Workers.

NORMAN R. F. MAIER. An Experimental Test of the Effect of Training on Discussion Leadership.

Book Reviews.

No. 3

HERBERT C. KELMAN. Attitude Change as a Function of Response Restriction.

ELÉONORE L. HERBERT and E. L. TRIST. The Institution of an Absent Leader by a Students' Discussion Group.

HAROLD B. GERARD. The Effect of Different Dimensions of Disagreement on the Communication Process in Small Groups.

NEHEMIAH JORDAN. Behavioral Forces that are a Function of Attitudes and of Cognitive Organizations.

Book Reviews.

30*s.* *per annum* or 8*s.* 6*d.* *per issue*

The Tavistock Institute of Human Relations
London, England.

The Research Center for Group Dynamics
Ann Arbor, Mich.

Subscriptions and orders to Managing Editor

TAVISTOCK PUBLICATIONS LTD

2, Beaumont Street, London, W.1

THE BRITISH JOURNAL OF PSYCHOLOGY

Edited by D. W. HARDING

Vol. XLIV. Part 3. **August, 1953** **12s. 6d. net**

OBITUARY NOTICE. David Katz.

C. A. MACE. Homeostasis, needs and values.

W. M. O'NEIL. Hypothetical terms and relations in psychological theorizing.

W. KENNETH RICHMOND. Educational measurement: its scope and limitations. A critique.

L. W. SHEARS. The dynamics of leadership in adolescent school groups.

F. H. GEORGE. 'Either-or' questions in series.

F. A. CHRENKO. Probit analysis of subjective reactions to thermal stimuli—a study of radiant panel heating in buildings.

G. ROBERT GRICE. Hunter's test of the absolute and relative theories of transposition.

IAN M. L. HUNTER. Reply to Professor Grice.

PUBLICATIONS RECENTLY RECEIVED.

The subscription price per volume, payable in advance, is 40s. net (post free)

Subscriptions may be sent to any bookseller or to the

CAMBRIDGE UNIVERSITY PRESS

BENTLEY HOUSE, EUSTON ROAD, LONDON, N.W.1

OCCUPATIONAL PSYCHOLOGY

Editor: ALEC RODGER

October, 1953 CONTENTS Volume 27, No. 4

Difficulties of Older People in Industry. By R. M. BELBIN.

Attitudes of Women Operatives to Promotion. By ELIZABETH LIVINGSTONE.

Some Effects of Joint Consultation on the Status and Role of the Supervisor. By H. CAMPBELL.

Seasonal Fluctuation in the Average Ability of Men Available for National Service. By SYLVIA ANTHONY.

The Mobility of Labour in Great Britain. By GEOFFREY THOMAS.

Book Reviews.

Other Books Received.

Index.

Annual Subscription 30 *shillings*

National Institute of Industrial Psychology

14, Welbeck Street, London, W.1

PRINTED IN ENGLAND BY HAZELL, WATSON & VINEY, LTD., LONDON AND AYLESBURY—8826-53

UNIVERSITY OF LONDON 11 DEC 1953 LIBRARY